PROJECT ORION

A Design Study of a System for Detecting Extrasolar Planets

NASA

NASA SP-436

PROJECT ORION

A Design Study of a System for Detecting Extrasolar Planets

David C. Black, Editor
Ames Research Center

National Aeronautics and
Space Administration

**Scientific and Technical
Information Branch
1980**

Cover: Orion Nebula XII
Photograph courtesy of Lick Observatory.

Ode to Apodization

Twinkle, twinkle, little star
Thirty parsecs from where we are.
Does your wobble through the sky
Mean a planet is nearby?
Or has the result come into being
Because of one arcsecond seeing?

—Raymond P. Vito

Library of Congress Cataloging in Publication Data
Main entry under title:

Project Orion.

(NASA SP ; 436)
Bibliography: p.
1. Project Orion. I. Black, David C. II. Series: United States. National Aeronautics and Space Administration. NASA SP ; 436.
QB602.9.P76 523.1'13 80-11728

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

TABLE OF CONTENTS

Conceptual drawing of Orion.

Stonehenge, Salisbury Plain, England
Photograph courtesy of Beulah P. Gossett.

PREFACE

The motivation behind Project Orion can perhaps be best understood by answering the question "Why is a search for other planetary systems important?" In large measure, the answer to that question is contained in two topics: the origin of the solar system and the existence of extraterrestrial intelligence – topics that seem at first glance unrelated.

Speculation concerning the origin of the solar system is as old as man himself. The "modern" era of such speculation is generally considered to have begun with Copernicus, who showed that Earth is not the center of the universe. The major cosmogonic hypotheses of this early period were those of Laplace and Kant. Although differing from each other in detail, their hypotheses were similar in that both men envisioned the planets as having formed from material that was spun off the Sun. These hypotheses were discarded a little over a century after they were advanced when it was realized that the Sun, which has about 99 percent of the mass in the solar system, has only about 2 percent of the angular momentum of the solar system. The principal reason these hypotheses fell into disfavor was that, at the time of this observational discovery, there was no known mechanism that would remove angular momentum from the Sun, thereby making it hard to understand why the planets have so much angular momentum relative to the Sun if they were originally part of the Sun.

The next generation of cosmogonic hypotheses was directed specifically at the problem of the angular momentum distribution in the solar system. Again, there were variations among the hypotheses, but they all tended to rely on singular or catastrophic events, such as a supernova or a stellar collision, to account for the solar system. Most of these "catastrophic" hypotheses, which were popular in the first few decades of this century, have subsequently been shown to be physically untenable or, at the very least, highly unlikely.

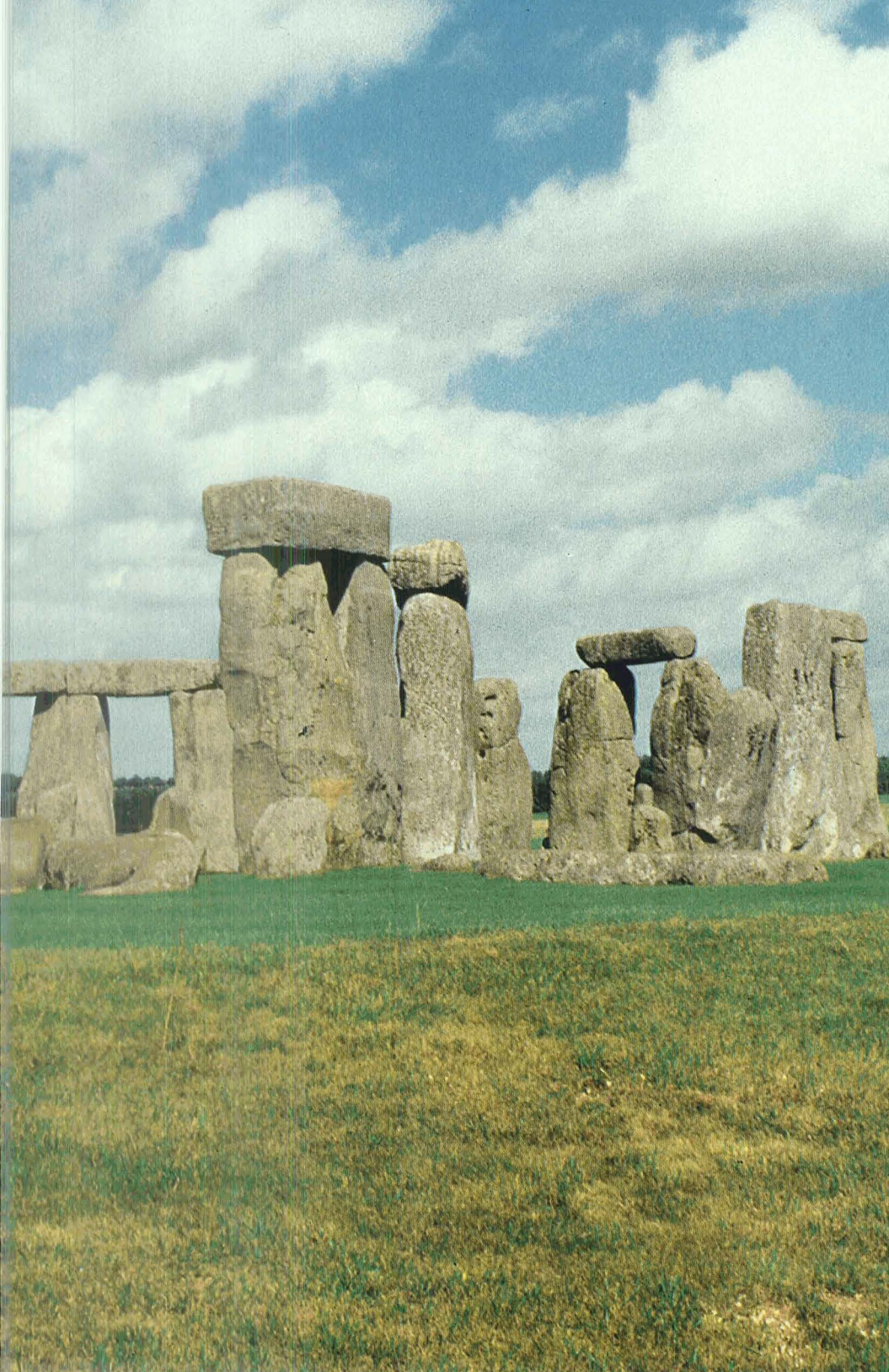
Most current hypotheses concerning the origin of the solar system are in many ways similar to the early hypotheses of Laplace and Kant. They envision the planets forming from a nebula of gas and dust which is thought to have surrounded the Sun early in its

history. This return to a "nebular" hypothesis has come about primarily for two reasons. First, unlike our counterparts of the previous century, we now know of several reasonable physical mechanisms that could have removed angular momentum from the Sun. Combining observational studies of the rotational velocities of stars on the main sequence (stars of various masses which, like the Sun, are deriving their luminous energy mainly from conversion of hydrogen into helium) with theoretical models of stellar structure, a consistent picture emerges which indicates that the Sun has very little angular momentum because most of it was removed by the solar wind, a plasma that continuously flows off the Sun's surface. The second reason for a return to "nebular" hypotheses is that the advent of high-speed electronic computers permits us to perform numerical experiments or modeling of the collapse of an interstellar cloud under the influence of its own gravity. It is generally felt that stars form by virtue of such a collapse process. Although these numerical experiments are not definitive, they strongly suggest that when a star is born it is likely to have a circumstellar nebula, and that conditions in such a nebula would be highly conducive to the formation of planetary bodies. An equivalent way of expressing present cosmogonic hypotheses is to say that planetary system formation seems to be a natural, if not inevitable, aspect of the star formation process. The important distinction between the catastrophic and nebular cosmogonic hypotheses is that, if the former is correct, planetary systems are the exception rather than the rule, whereas if the latter is correct, planetary systems are the rule. A systematic study of the frequency of occurrence of planetary systems would thus provide a valuable observational check on present theories of star formation.

The possible existence of extraterrestrial intelligence (ETI) has long piqued man's curiosity. This curiosity has been the basis for a number of science fiction efforts. A graphic demonstration of this fascination with ETI is Orson Welles' radio dramatization in 1939 of the work by H. G. Wells entitled, "War of the Worlds." Until relatively recently, ETI had been the plaything of science fiction, largely ignored as a subject of scientific inquiry. However, Cocconi and Morrison in 1959 took the first major step in changing the attitude of the scientific community toward the question of ETI. They pointed out that there is a natural signpost in the electromagnetic spectrum that would be known to any advanced civilization, and that

such civilizations might send radio signals at or near the frequency of this natural marker. This signpost is the 21-cm-wavelength radiation arising from a hyperfine transition in atomic hydrogen, the most abundant element in the universe. Shortly after the paper by Cocconi and Morrison, Frank Drake conducted a search, known as Project Ozma, for such signals. Drake's search was unsuccessful, but its importance cannot be overlooked as it was the first serious attempt at detecting ETI signals. The relevance of a SETI (search for ETI) effort to a search for other planetary systems lies in the fact that the only known intelligent life form, namely ourselves, developed and was nurtured on a planet. If planets are required for the existence of ETI, knowledge of the frequency of occurrence of planetary systems is clearly highly desirable.

A systematic search for other planetary systems would thus reveal whether there is justification in arguing that a natural, perhaps even causal, relation exists between the phenomenon of star formation, which has occurred some 10^{11} times in the galaxy, and the existence of ETI. The detection of other planetary systems is difficult; present observational techniques and instrumentation are at best marginal in terms of their ability to carry out such a search. The purpose of this design study was to apply modern technology to the problem in the form of specific design concepts for systems that could successfully mount a search for other planetary systems. The Earth is an object of exquisite beauty, and to the extent that the material on the ensuing pages is instrumental in the discovery of another such object, it will have rendered mankind an invaluable service.



1. INTRODUCTION

Man's fascination with the heavens is as old as recorded history. Ancient astronomers carefully observed the relative position of stars, as well as the time of year when a particular star or group of stars was visible in the night sky. Two characteristics of the motion of stars were apparent to those early observers. First, the general movement of stars from one night to the next is a slow westward drift, with stars periodically disappearing from view for long intervals of time. Second, there is an overall unison in this motion. Each star moves so as to keep its position fixed relative to its neighboring stars. It is against this background of periodic and uniform motion that the story of planetary detection began.

DISCOVERY OF OUR PLANETARY SYSTEM

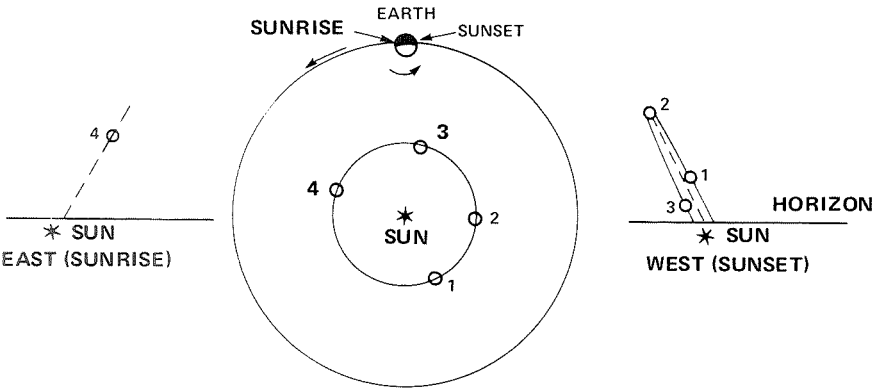
Early in man's studies of the night sky, it was noticed that five "stars" do not participate in the harmony of motion shared by most celestial objects. The movement of these five planets (from "planētēs," Greek for "wandering") baffled man for thousands of years, during which time they played central roles in much of formalized religious and mythological thought.

What is it that made the planets so special? Unlike other stars, all of the planets drift nightly from west to east relative to the stars. There were also important distinctions among the planets. Two of them appear either as evening or morning "stars," that is, they are either visible shortly after sunset or before sunrise. However, they do not move from the evening to morning sky by passing overhead. Once they appear in the evening sky, they rise higher and higher in the sky on successive nights until they reach a maximum elevation above the horizon. They then begin to recede toward sunset on subsequent evenings until they are lost in its glare. After an interval of many weeks, these two planets reappear just before sunrise and repeat the behavior observed in the evening sky: a gradual rise to a

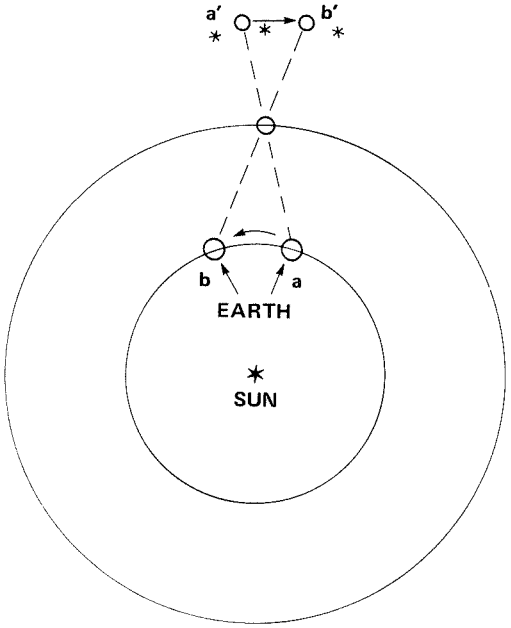
maximum elevation many nights after their appearance in the morning sky, followed by a return to sunrise. The remaining three planets do pass continuously from evening to morning sky, but roughly once a year they appear to move backward in the sky. The extent of this reversed or retrograde motion is different for each planet.

In 1543, Nicolaus Copernicus published his book, "The Revolutions of the Celestial Spheres," providing the explanation of the apparent motion of the five nomads. Copernicus stated that the Earth is not the center of the universe, rather that it and the five wanderers are all planets revolving around the Sun. Figure 1(a) shows how Copernicus' model explains the motion of the two evening-morning stars. Note that the planet nearest the Sun is visible in the evening sky at position 1, moves to a maximum separation from the Sun at position 2, is visible near sunset at position 3, and is seen in the morning sky before sunrise at position 4. The explanation of the nearly annual retrograde motion of the other three planets is shown in figure 1(b). Over most of Earth's orbit, the outer planet appears to move eastward in the sky relative to the distant stars. However, as Earth overtakes the outer planet (positions a and b in the figure), the latter *appears* to move backward (from a′ to b′) in relation to the distant stars. The extent of retrograde motion is largest for the planet closest to Earth's orbit. Copernicus' "discovery" of our planetary system forever altered man's concept of himself and his relationship to the universe as a whole.

The solar system family of planets at the time of Kepler and Galileo consisted of Mercury, Venus, Earth, Mars, Jupiter, and Saturn. The next addition to the family occurred in 1791 when the planet Uranus was discovered by Sir William Herschel. Uranus was carefully observed by Herschel and others for many years, at which time it became apparent that something was wrong. The planet was not following its predicted orbit. Leverrier of France and Adams of Britain analyzed the planet's motion and independently arrived at the conclusion that the perturbation in Uranus' orbit was due to another, more distant planet. Their calculations and predictions were generally ignored by their fellow astronomers. However, the planet Neptune was discovered in 1846 by Galle of the Berlin Observatory. Galle used Leverrier's work and found the planet within 1° of the predicted position. The discovery of Neptune is extremely important because it was the first planet detected by combining the science of



(a) Apparent motion of one of the "evening-morning stars."



(b) Apparent motion of planets with orbits larger than Earth's.

Figure 1.– *Schematic representation of the relative motions of the planets.*

astronomy and the discipline of mathematics. The last planetary member of the solar system to be discovered was Pluto. The existence of Pluto was postulated in 1914 by Galliot and Lowell to explain perturbations in the orbits of Neptune and Uranus. It was discovered in 1930 by Tombaugh working at the Lowell Observatory.

EFFORTS TO DETECT OTHER PLANETARY SYSTEMS

Scientific and technological developments in the late nineteenth and early twentieth centuries made it possible to measure the separation between two photographically recorded star images with accuracies of a few thousandths of a millimeter. As a consequence of this accuracy, it became possible to measure stellar angular separations as small as a few hundredths of an arcsecond (an arcsecond is an angle equal to 1/3600 of a degree or 1/206,265 of a radian).

The first unseen object to be detected outside the solar system was not a planet, but a new type of star. Astronomers had long puzzled over small variations in the motion of the star Sirius. Instead of moving along a straight line, Sirius wobbled from one side to the other of its predicted course (fig. 2). F. W. Bessel calculated that the

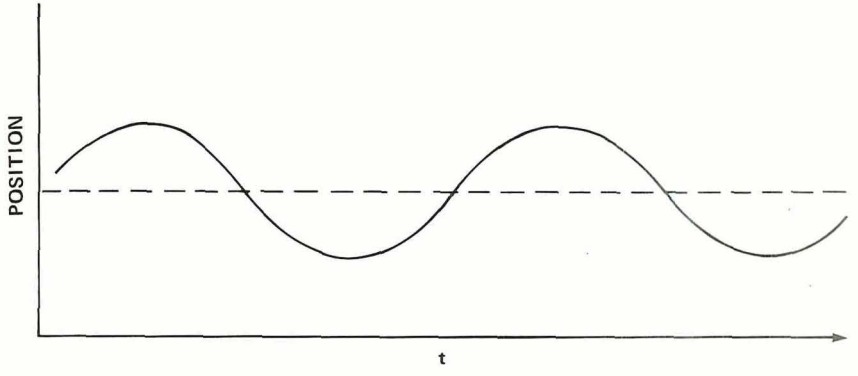


Figure 2.– *Schematic representation of the position of a star, as a function of time* t, *with a dark companion. Dashed line shows the position (a constant value) if the star has no companion.*

perturbations in the motion of Sirius were due to an unseen companion of considerable mass. Although Bessel was able to predict the position of the companion, it remained undiscovered until 1864. Later studies show that the companion's mass is approximately equal to the mass of the Sun.

The perturbation in Sirius' motion is only about 4 arcsec. If the companion had the mass of Jupiter rather than that of the Sun, the perturbation would have been proportionately smaller. Such a small perturbation is at the limit of the measurement capability of present telescopes. However, there are a number of stars less massive than Sirius, a few of which are also closer to the solar system. Peter van de Kamp of the Sproul Observatory, realizing that perturbations in the motion of these nearby, less massive stars might be detectable, initiated the first systematic search for other planetary systems.

The first results with the Sproul Observatory's 24-in. telescope were presented in 1943 by K. A. Strand. The results indicated that one of the two small stars in the binary star system, known as 61 Cygni, had a dark, or unseen, companion. The 61 Cygni study was soon followed by other tentative discoveries of planetary companions to small stars with such strange-sounding names as BD + 43° 4305, Epsilon Eridani, Lalande 21185, and Barnard's star (see table 1). But there were bothersome similarities in these results. The perturbations were always at the limit of the error of the studies. Also, the orbital periods of the proposed planets discovered with the Sproul telescope were all very nearly multiples of 8 years. The 61 Cygni study is an exception; it is only partially based on data from the Sproul telescope. It is also worrisome that the eccentricities

TABLE 1.– SUSPECTED PLANETARY SYSTEMS

Name of system	Distance (light yr)	Planet masses (Jupiters)	Periods (yr)
Barnard's Star	5.9	1.1, 0.8	26, 12
Lalande 21185	8.2	10	8.0
Epsilon Eridani	10.8	6–50	25
61 Cygni	11.0	8	4.8
BD + 43° 4305	16.9	10–30	28.5

of the calculated planetary orbits are all large, averaging above 0.5. By comparison, the eccentricities of planetary orbits in the solar system are generally below 0.1 (Mercury and Pluto being the only exceptions).

Perhaps the best known study of this kind is that of Barnard's star. This is the only case in which the proposed dark companion has a mass comparable to that of a planet in the solar system (refs. 1, 2). A more recent study (ref. 3) has shown that the perturbation evidenced in the Sproul plates does not appear on plates taken with the 20-in. telescope of the Van Vleck Observatory, nor does it appear on plates taken with the 30-in. Thaw telescope at the Allegheny Observatory. Hershey (ref. 4) showed that there are epochs of "discontinuity" in the Sproul data which coincide with times when the lens of the telescope was adjusted. In light of demonstrations that the suspected perturbations were instrumental in origin, it may be said that *there is no unequivocal evidence for the existence of planets outside the solar system.*

PROJECT ORION

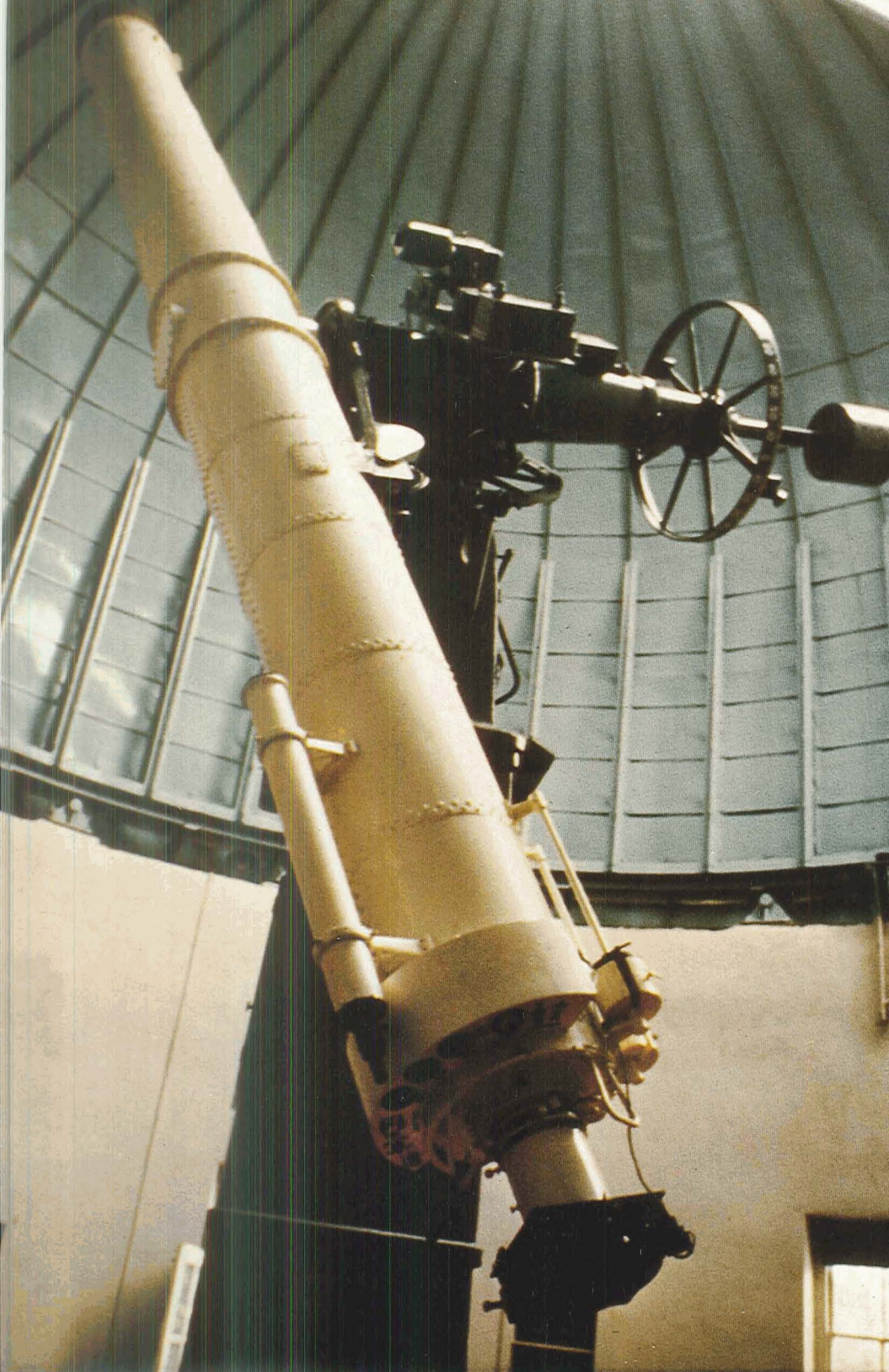
With one exception, the telescopes presently used in astrometry, that branch of astronomy concerned with the position and motion of the stars, are about 60 years old, and only two were designed for high-precision astrometric observations. The rest have been converted to photographic use by addition of a colored filter in conjunction with yellow-sensitive photographic emulsions. None of the existing instruments are designed to meet the requirements of ultrahigh precision and long-term stability necessary for successful detection of very small perturbations.

An important aspect of a search for other planetary systems concerns the extent of the search: Is it to include a handful of nearby stars in hope of discovery, or is it to include a sufficiently large number of stars to provide a basis for meaningful statistics on the existence of other planetary systems? Discovery of another planetary system would be a significant event, but the fundamental questions are "What is the frequency of occurrence of planetary systems?" and "Do all types of stars have planetary companions?"

Present telescopes may succeed in discovering a planet revolving around a nearby star, but they will leave the fundamental questions unanswered. A new instrument (or instruments) is needed. Technology has advanced considerably since the last large refracting telescope was made, and it is reasonable to suppose that, through application of modern technology, a telescope could be built that would be far more accurate than existing instruments. Project Orion is a first step toward the goal of developing such a telescope, perhaps more importantly a first step toward a comprehensive program to search for extrasolar planets.

The classic technique for searching for extrasolar planets is that of precise astrometric studies. Because the field of astrometry has not enjoyed an infusion of modern technology to the extent that its sister subdisciplines within astronomy have, it was decided that the major emphasis of Project Orion would be to develop a design concept for a ground-based astrometric telescope that could, in principle, significantly increase the potential accuracy of astrometric observations. In the spring of 1976, two workshops were held to examine the state of current techniques and instrumentation in the context of detecting extrasolar planets (unpublished report to the Office of Space Science of the NASA, 1976). Among the techniques that seemed potentially capable of detecting extrasolar planets, two emerged as feasible alternatives to existing techniques. Both involved detection of light from extrasolar planets, in contrast to astrometric studies in which the presence of a planet is inferred from its effect on an observed star. A portion of Project Orion involved studies of these two techniques whereby extrasolar planetary radiation – in one case, intrinsic thermal radiation, in another case, reflected visual radiation – might be detected.

61-cm Sproul Refractor
Photograph courtesy of Sproul Observatory, Swarthmore College, Swarthmore, Pennsylvania.



2. TOWARD DESIGN CONCEPTS

It is useful to examine the magnitude of the planetary detection problem and the fundamental limitations of either a physical or technological nature, particularly as such considerations significantly affect systems design concepts. This chapter outlines some of the key factors involved in Project Orion as it progressed from a definition of the problem toward the formulation of the design concepts discussed later.

REMARKS CONCERNING THE TERM "PLANET"

The definition of "planet" used by ancient astronomers centered on the erratic apparent motion of certain celestial bodies. It is perhaps surprising that there is no generally accepted modern definition of "planet." Questions arise as to whether the definition should be only in terms of intrinsic properties of a body (e.g., mass) or whether it should also include properties related to orbital motion (e.g., that the orbit be nearly circular). The view taken here is that the term "planet" refers to any object whose mass is comparable to or less than the mass of Jupiter ($\sim 1.9 \times 10^{27}$ kg). Adoption of this operational definition does *not* imply that bodies which are more massive than Jupiter are not planets; the upper limit of the mass of a planet is set by the lowest mass consistent with the definition of a star ($\sim 0.06 M_{\odot}$) (subscript $\odot$ denotes the Sun, so that $M_{\odot}$ = mass of the Sun, or 1.998×10^{30} kg). The choice of this definition of planet arises directly from the challenge addressed in this study, namely, to design instrumentation that would permit unambiguous detection of objects with the mass of Jupiter. Such instrumentation would detect more massive objects with relative ease.

DETECTION PROBLEM – ASTROPHYSICAL ASPECTS

The three techniques for detecting extrasolar planets considered here are: (1) astrometric searches for perturbations in stellar motion, (2) infrared searches for intrinsic thermal extrasolar planetary radiation, and (3) optical searches for reflected visual stellar radiation. For reasons indicated later, technique 1 is referred to as an indirect detection technique, whereas techniques 2 and 3 are referred to as direct detection techniques. It will be useful to briefly discuss the physical basis for effects arising from extrasolar planets which give rise to phenomena that may be studied by one or more of the techniques listed above. This discussion will also serve to indicate both the informational content (i.e., what might be learned concerning a detected planet) of each technique and the order of magnitude of the detection problem.

Observables and Governing Equations

What are the observables related to each of the three techniques considered here, and what are the relationships between those observables and characteristics of a potential extrasolar planet? Imagine a rather simple planetary system consisting of a star and a single planetary companion (fig. 3). The mass, diameter, and effective temperature of the star are denoted by M_*, d_*, and T_*, respectively, whereas the corresponding parameters for the planetary companion are denoted by M_p, d_p, and T_p. The barycenter (center of mass) of this ersatz planetary system is located at distances R_* and R_p from the centers of mass of the star and planet, respectively, where $M_p R_p = M_* R_*$ with

$$R_* = \left(\frac{M_p}{M_* + M_p} \right) R \tag{1}$$

and

$$R_p = \left(\frac{M_*}{M_* + M_p} \right) R \tag{2}$$

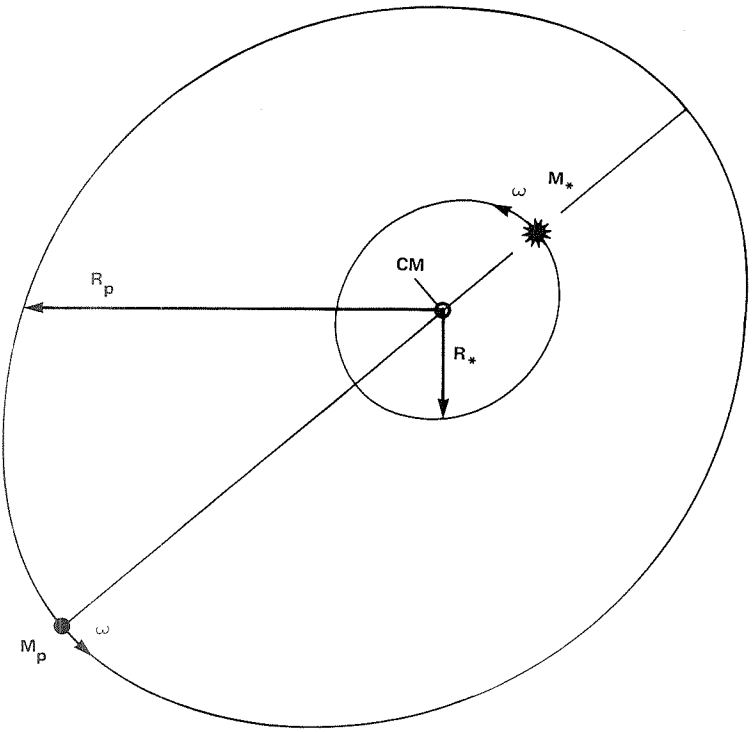


Figure 3.– *Schematic representation of a simple planetary system (indicated parameters are defined in the text).*

The quantity $R = R_* + R_p$ is the semimajor axis of the planet's orbit about the star, taken here to be circular to simplify the discussion. The three techniques are discussed in the order listed above and in the context of detecting this simple planetary system.

A detailed discussion of astrometry will not be given here, but interested readers can consult any number of excellent books on the subject (e.g., ref. 5). If the system under study is an isolated (no companions) star, the barycenter lies at the center of mass of the star. If, however, the star has a companion (planet?), the system barycenter is displaced from the star's center of mass by a distance R_* (eq. (1)). In this latter case, the star and companion will revolve about the barycenter with a period determined by M_*, M_p, and R. Precise observations of the star could, in principle, reveal that its motion departed from rectilinear motion.

The amplitude of this nonlinear motion is R_*. However, R_* is not the directly observed quantity. The observed quantity is essentially an angle θ, where

$$\theta = \arctan\left(\frac{R_*}{D}\right) = \arctan\left(\frac{M_p}{M_*} \times \frac{R_p}{D}\right) \tag{3}$$

where D is the distance from Earth to the star under observation. In general, $M_p R_p << M_* D$, so that

$$\theta \cong 9.8 \times 10^{-4} \left(C_1 \frac{M_p}{D}\right) \text{ arcsec} \tag{4}$$

where M_p and M_* are, respectively, in units of Jovian and solar masses, $C_1 \equiv R_p/M_*$, R_p is in astronomical units (AU) (1 AU = 1.5×10^{11} m, the mean distance from Earth to the Sun), and D is in parsecs (1 pc = 3×10^{16} m). Representative values of θ as a function of D are shown in figures 4 and 5, respectively, for $M_p = 1$ and 0.003 (an Earth mass planet). The technique of detecting the presence of extrasolar planets by astrometric observations is referred to as indirect detection because the presence of a planet is deduced from observations of a star, *not* of the planet.

Planetary companions to stars are "sources" of electromagnetic radiation. This radiation can be characterized as either thermal or reflected. A planet will radiate at some temperature which is determined by a balance between the rate at which the planet receives energy from both internal (e.g., radioactive) and external (e.g., its parent star) sources of energy, and the rate at which the planet loses energy by radiation. This thermal radiation is most pronounced in the infrared region of the spectrum because planets are relatively cool objects. A planet can also reflect radiation from its parent star, the amount of reflected radiation depending both on the size of the planet and the nature of the reflecting medium. This reflected component of a planet's radiation spectrum is generally strongest at those wavelengths where the central star emits most of its radiation, namely, the visual portion of the spectrum. The total energy flux

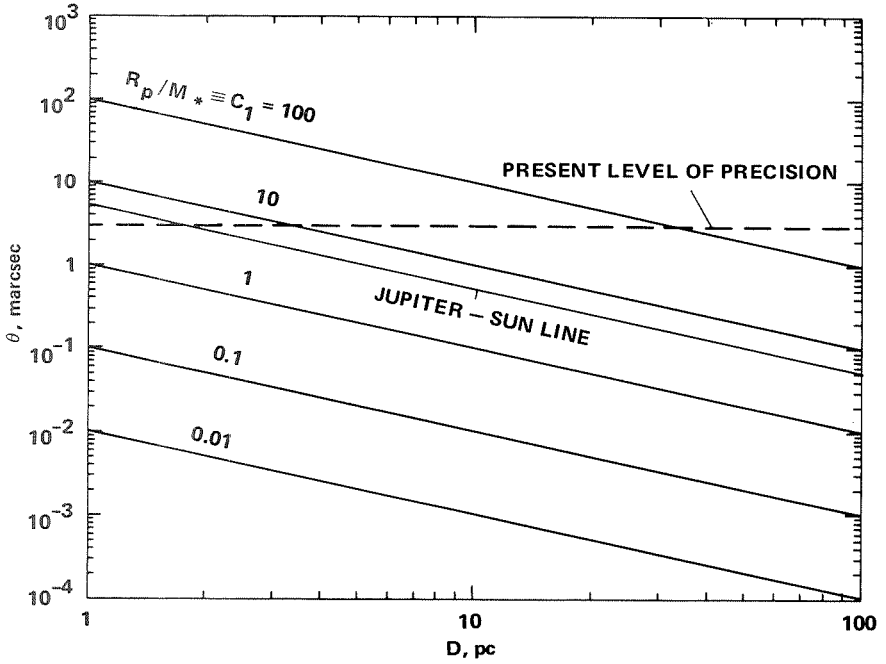


Figure 4.– *Maximum angular displacement* (θ) *of a star due to a Jovian mass companion shown as a function of its distance* (D) *from an observer. Present level of observational accuracy indicated by dashed line; various slanting lines refer to differing values of* R_p/M_*.

(per unit frequency interval) from a planet is the sum of its intrinsic thermal flux, $\Phi_{TH}(\nu)$, and reflected stellar flux, $\Phi_{REF}(\nu)$:

$$\Phi_p(\nu) = \Phi_{TH}(\nu) + \Phi_{REF}(\nu) \tag{5}$$

If $\Phi_*(\nu)$ represents the radiant energy flux leaving the surface of the star depicted in figure 3, only a fraction f_p of that flux will fall on the planet, with

$$f_p = \left(\frac{d_p}{4R}\right)^2 \tag{6}$$

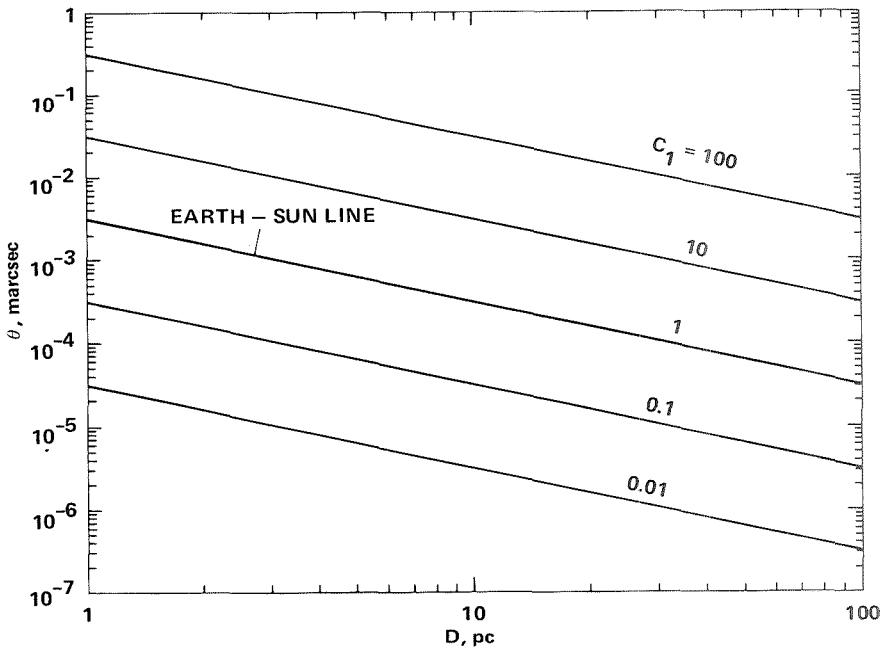


Figure 5.– *Maximum angular displacement* (θ) *of a star due to an Earth mass dark companion shown as a function of its distance* (D) *from an observer. Present level of observational accuracy indicated by dashed line; various slanting lines refer to differing values of* R_p/M_*.

In general, only a fraction of the radiation incident upon a planet will be reflected (that fraction denoted by η):

$$\Phi_{REF}(\nu) = \eta f_p \Phi_*(\nu) \tag{7}$$

The parameter η is usually a function of frequency, but that complication is ignored here. Assuming that the intrinsic thermal flux from the planet is that of a blackbody of temperature T_p gives

$$\Phi_{TH}(\nu) = \frac{2\pi^2 d_p{}^2}{c^2} \cdot \frac{h\nu^3}{(e^{h\nu/kT_p} - 1)} \tag{8}$$

where c, h, and k are, respectively, the speed of light in vacuum, Planck's constant, and Boltzmann's constant. Combining equations (5), (7), and (8) gives

$$\Phi_p(\nu) = C_1 {d_p}^2 \left[C_2(\nu, T_p) + \frac{\eta {d_*}^2}{16 {R_p}^2} C_2(\nu, T_*) \right] \nu^3 \tag{9}$$

where the constant $C_1 = 2\pi^2 h/c^2 = 1.463 \times 10^{-49}$ (in mks units) and

$$C_2(\nu, T) = \left[\exp\left(\frac{h\nu}{kT}\right) - 1 \right]^{-1}$$

The number of photons emitted per second per unit frequency interval, $N_p(\nu)$, is obtained by dividing $\Phi_p(\nu)$ by the energy per photon, $h\nu$. Equation (9) gives the flux at the planet's surface. Of more interest from the detection standpoint is the energy $E_{DET}(\nu)$ which arrives at a detector located on or near Earth. If one ignores effects such as limb darkening, $E_{DET}(\nu)$ is given approximately by

$$E_{DET}(\nu) = \left[\frac{A \cdot B(\nu)}{4\pi D^2} \right] \Phi_p(\nu) \tag{10}$$

where A and B are, respectively, the area and the frequency bandwidth of the detector, and D is the distance defined previously. The number of photons, $N_{DET}(\nu)$, is given by

$$N_{DET}(\nu) = \frac{E_{DET}(\nu)}{h\nu} \tag{11}$$

for $B << \nu$. Figure 6 shows $\Phi(\nu)$ for the Sun and $\Phi(\nu)$ for Jupiter, assuming that $\eta(d_p/4R_p)^2 = 2 \times 10^{-9}$.

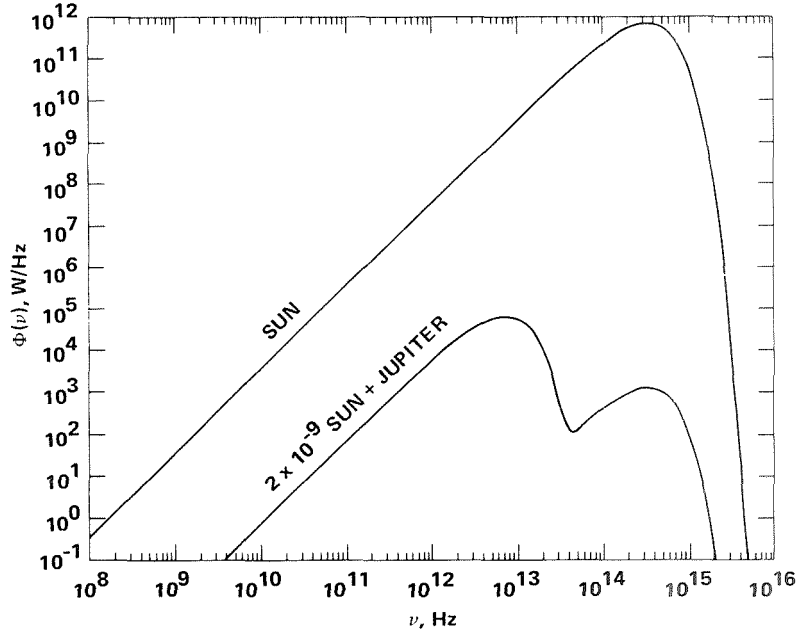


Figure 6.– *A plot of the power emitted per unit frequency interval as a function of frequency. Data are shown for the Sun (upper curve) and for Jupiter. The hump in the Jupiter curve at low frequencies (~10^{13} Hz) is due to intrinsic thermal radiation, whereas the hump at higher frequencies (~3×10^{14} Hz) is due to reflected sunlight.*

One other quantity of interest is the ratio $H(\nu)$ of planetary radiation to stellar radiation of frequency ν, namely,

$$H(\nu) = \frac{\Phi_p(\nu)}{\Phi_*(\nu)} = \frac{N_p(\nu)}{N_*(\nu)}$$

$$= \frac{d_p}{d_*}^2 \left[\frac{C_2(\nu, T_p)}{C_2(\nu, T_*)} + \frac{\eta {d_*}^2}{16 {R_p}^2}\right] \qquad (12)$$

where $\Phi_*(\nu)$ is given by equation (8) with d_* and T_* substituted, respectively, for d_p and T_p. At high frequencies ($h\nu >> kT_*$),

$H(\nu)$ approaches the limiting value $\eta(d_p/4R_p)^2$, which is independent of stellar properties and of ν. At low frequencies ($h\nu << kT_p$), $H(\nu)$ is independent of ν, but depends on both stellar and planetary properties, namely,

$$H(\nu << kT_p/h) \sim \frac{{d_p}^2 T_p}{{d_*}^2 T_*}$$

where $e^{h\nu/kT} - 1 \cong h\nu/kT$ and $\eta(d_*/4R_p)^2 << T_p/T_*$. The techniques of detecting the presence of extrasolar planets by observations of planetary radiation, either intrinsic thermal or reflected visual radiation, are referred to as direct detection techniques because detection arises from a direct observation of the planet.

Equations (4) and (10)–(12) provide the analytical basis for estimating the order of magnitude of observable parameters for assumed extrasolar planetary systems and for determining what each of the three techniques considered here might provide by way of information concerning these planetary systems.

A Standard System

In order to make subsequent discussion and numerical examples specific, it is useful to define a "standard planetary system" (SPS). The basic free parameters that define an SPS are those depicted in connection with figure 3, as well as the number of planets. The standard chosen here is that of a spectral type G2 main sequence star (e.g., the Sun) around which revolves a single planet. The mass, dimensions, and temperature of that planet are taken to be identical to those of Jupiter. The planet's orbit is taken to be circular with a radius of 5 AU. The principal characteristics of the SPS are summarized in table 2. The equations given earlier indicate how observable parameters scale with different assumptions concerning the characteristics of the star, planet, and orbit.

TABLE 2.– CHARACTERISTICS OF STANDARD PLANETARY SYSTEM

T_* (K)	d_* (m)	M_* (kg)	T_p (K)	d_p (m)	M_p (kg)	R_p (m)
5800	1.39×10^{9}	1.99×10^{30}	128	1.43×10^{8}	1.90×10^{27}	7.5×10^{11}

Numerical Examples and Information Content

The order of magnitude for observables relating to the SPS are obtained here for two assumed values of the distance parameter D, namely, $D = 10$ and 30 pc. Equation (4) indicates that the magnitude of the angular perturbation at these distances is

$$\theta(D = 10 \text{ pc}) = 9.8\times10^{-4} \times 0.5 = 4.9\times10^{-4} \text{ arcsec}$$

and

$$\theta(D = 30 \text{ pc}) = 9.8\times10^{-4} \times 0.16 = 1.6\times10^{-4} \text{ arcsec}$$

By comparison, the highest accuracy obtained with current astrometric observations is about 3×10^{-3} arcsec, and typical performance is more like 5×10^{-3} arcsec. These accuracy figures pertain to so-called yearly mean normal points; that is, they represent the results of many observations per year with the accuracy of the mean increased over the accuracy of individual measurements in proportion to the square root of the number of observations per year. Note that the performance necessary to detect the SPS at $D = 10$ pc exceeds that of present astrometric facilities by a factor of 6 and by a factor of about 20 if it is required that the formal error be no greater than 1/3 the maximum values of θ.

Equation (10) gives the fraction of energy flux from an extrasolar planet which arrives at Earth. It is convenient to estimate the photon flux rather than the energy flux, as photon noise is an important parameter in the detection of radiation. This noise varies inversely as the square root of the number of photons. The frequency ν_m of maximum photon emission per unit *wavelength* interval from a blackbody at temperature T is given by $\nu_m = 8.175\times10^{10}$ T (Hz). Thus, ν_m for the star in the SPS is $\nu_m{}^* = 4.74\times10^{14}$ Hz ($\lambda = 6.33\times10^{-7}$ m), whereas ν_m for the planet in the SPS is $\nu_m = 1.05\times10^{13}$ Hz ($\lambda = 2.87\times10^{-5}$ m). Evaluating the photon flux equivalent of equation (10) at these two frequencies and at distances of $D = 10$ and 30 pc yields the results given in table 3. Note that the fluxes given in table 3 are per unit *wavelength* interval,

not per unit *frequency* interval. The detector bandwidth and telescope aperture used to determine the values given in table 3 are, respectively, $B = 0.1\ \nu$ (Hz) and $A = 1\ m^2$.

The values of N_{IR} and N_V given in table 3 indicate that the standard planet would not be readily detectable in the visible portion of the spectrum beyond distances of about 4 pc ($N_V \sim 1$/sec) with $A = 1\ m^2$ and $B = 0.1\ \nu$, but could be detected at even greater distances if either more collecting area or more light (larger B) were used in the observation. The important point is not that a specific planetary system is or is not detectable using a specific set of parameters, but that both visible and especially infrared radiation from extrasolar planets revolving around stars located at reasonable distances (~10–30 pc) from Earth gives rise to a detectable number of photons. The question of detectors and noise sources for these measurements is discussed elsewhere.

The essential aspects of information content for direct detection are contained in equation (10). The net energy flux on a detector, $E_{DET}(\nu)$, is essentially the measured quantity. The frequency ν and fractional bandpass $\Delta\nu/\nu$ are also known quantities. Assuming that distance D to the star under observation is known independently, equation (10) may be recast in the form:

$$M(\lambda) = d_p{}^2 \left[C_2(\lambda, T_p) + \eta \left(\frac{d_*}{4R_p} \right)^2 C_2(\lambda, T_*) \right] \qquad (13)$$

where

$$M(\lambda) = \left[\frac{4\pi D^2 \lambda^5}{C^4 C_1 A B(\lambda)} \right] E_{DET}(\lambda)$$

TABLE 3.– PHOTON FLUX/UNIT WAVELENGTH INTERVAL FROM SPS PLANET[a]

Distance (pc)	$N_{IR}(\nu = 1.05\times10^{13}$ Hz)	$N_V(\nu = 4.7\times10^{14}$ Hz)
10	~9	~0.16
30	~1	~0.018

[a] $A = 1\ m^2$; $\Delta\nu/\nu = \Delta\lambda/\lambda = 0.1$.

and the variable λ has been used rather than

$$\nu[\nu = c/\lambda, B(\lambda) = (\lambda^2/c)B(\nu)]$$

In the infrared portion of the spectrum,

$$C_2(\lambda, T_p) > \eta\left(\frac{d_*}{4R_p}\right)^2 C_2(\lambda, T_*) \tag{14}$$

so that

$$M(\lambda) \cong \frac{d_p^2}{e\left(\frac{hc}{\lambda k T_p}\right) - 1}$$

If observations are made at two infrared wavelengths, the temperature T_p can be determined, thereby making it possible to determine the size of the planet, d_p, from the measured parameter $M(\lambda)$. Further, the orbital period can also be determined from the observations. If the mass of the star and the orbital period are known, the value of R_p can be calculated from Kepler's third law. The usefulness of knowledge concerning R_p is that a comparison of T_p with temperature T_{BS} of a perfectly conducting black sphere ($T_{BS} = 0.354\ T_*\ (d_*/R_p)^{1/2}$) could provide an indication of atmospheric effects on the planet.

In the visual portion of the spectrum, the inequality given in equation (14) is reversed so that

$$M(\lambda) \cong \frac{\eta d_p^2 d_*^2}{16 R_p^2} \cdot \frac{1}{e^{hc/\lambda k T_*} - 1}$$

Values for T_* and d_* are known approximately from stellar evolution theory and spectral observations of stars. As in the infrared, the orbital period and hence R_p can be determined. However, this

knowledge only places constraints on the product ηd_p^2; there is no way to determine independently either η or d_p. It may be possible to make multicolor observations that could provide useful clues regarding the nature of the reflecting medium. For example, the Earth appears blue relative to solar light because of the selective reflection of shorter wavelengths by Earth's atmosphere.

The defining equation for astrometric studies is equation (4), with the measured quantity being directly related to θ. The mass of the star (M_*) under study can be estimated independently, and the distance to the star (D) can be measured. Thus, the product $M_p R_p$ can be related to measured or known quantities. The orbital period of the planet yields the value of R_p (given M_*), so that the mass of the planet can be determined. Although it is not clear from the somewhat simplified equations presented here, astrometric studies would also reveal much concerning the nature of the planet's orbit (e.g., its eccentricity and relative orientation to the line of sight).

Clearly, comprehensive studies to search for extrasolar planetary systems could, if successful, reveal something more about our neighbors in space other than their existence. The various kinds of information that can, in principle, be obtained are listed in table 4.

TABLE 4.– INFORMATION CONTENT OF DETECTION TECHNIQUES

Information about planet	Detection technique	
	Direct	Indirect
Orbital period	*	*
Orbital structure (inclination, etc.)		*
Temperature	*	
Atmospheric composition	*	
Mass		*

DETECTION PROBLEM – TERRESTRIAL ASPECTS

Having outlined the physical basis underlying the detection techniques considered here, it is useful to inquire as to limitations placed on the detection problem, and thereby on possible design concepts, by Earth's atmosphere.

Direct Detection Techniques

Earth's atmosphere presents many difficulties for any method of extrasolar planetary detection. The atmosphere is not quiescent, rather it is often highly turbulent, and it is generally disturbed by pressure waves. Light from a point source in space, such as a star, enters the atmosphere as a bundle of nearly parallel rays. However, due to atmospheric effects, these rays do not remain parallel. The principal atmospheric effect arises because all light rays do not traverse identical portions of the atmosphere. The net result is that the image of the star as seen at the bottom of the atmosphere is "smeared out" – part of the reason why stars do not appear as true points of light. This smearing out process is not uniform. As pressure waves propagate through the atmosphere, they may affect a given light ray differently than they affect neighboring light rays. This causes the apparent image to flicker or, in more familiar language, twinkle. The twinkling of stars is a consequence of their point source nature. A source sufficiently close so as not to be a point, such as a planet in the solar system, does not appear to twinkle because the twinkle arising from a particular portion of the object tends to cancel the twinkle from another portion of the object, so the eye perceives a fairly constant pattern.

The atmospheric effect discussed above is known as "seeing" and it is generally expressed in terms of the minimum size of a stellar image as seen through a telescope. Seeing at ground-based observatories typically ranges from 1.5 to 2.5 arcsec, but may occasionally be as good as ~0.5 arcsec at exceptional sites. Since the maximum angular separation in the standard system is only 0.5 arcsec, it is clear

that any direct imaging cannot be done from the ground; it must be done in space.[1]

Astrometric Observations

Astrometric techniques can be divided into those of absolute and relative astrometry. In absolute astrometry, stellar positions on the sky are measured relative to a system of coordinates defined by the direction of Earth's axis of rotation and the vertical at the place of observation. As in most other fields, absolute measurements are generally much less accurate than are relative measurements. In relative astrometry, the position of a star is measured relative to positions of other stars situated in angular proximity to it on the sky. Relative astrometry, with its inherently higher precision, is a more suitable approach than absolute astrometry for detecting planetary-induced wobble in the motion of a star.

The errors of measurements in relative astrometry can be traced to the following four sources:

1. Inaccuracy of devices used to measure the relative positions of star images formed by a telescope
2. Changes in relative positions of star images caused by instabilities of the optical components of a telescope
3. Effects of Earth's atmosphere
4. Effects intrinsic to the observed stars, such as variable color or brightness of reference stars

(Items 1 and 2 are discussed later; item 4 involves topics that fall outside the scope of this report, but are currently under investigation by some of the Project Orion personnel; and item 3 is discussed next.)

---

[1] Recent technological advances in the area of "active optics" or "rubber mirror telescopes" appear to make it possible to circumvent, to a large extent, "seeing" effects. However, there still does not appear to be any way, for ground-based observations, around the more fundamental difficulty of detecting a very dim object (planet) next to a very bright object (star). Therefore, the conclusion that direct detection of reflected stellar or intrinsic, thermal planetary radiation must be done from space is likely to remain valid. Further studies on this point are needed.

The angular separation between two stars can be split into two perpendicular components: a vertical component directed toward the zenith and a horizontal component. Measuring the vertical component requires knowledge of refraction in the atmosphere, which depends on the color of a star and on atmospheric pressure and temperature. The horizontal component, which is on the average less dependent on these factors, can be measured with considerably higher accuracy than the vertical component. Differences in both right ascension and declination of two stars can be obtained from measurements of the horizontal component of their angular separation if the observations are made at two azimuths, for example, at $A = \pm 30°$. The following discussion will therefore be limited to measurements of the horizontal component.

The analysis given here is based on Reiger's model (ref. 6) of atmospheric turbulence. This model successfully explains various observed properties of scintillation of starlight (ref. 7) and the average size of excursions of star images (appendix II in ref. 6). Reiger's theory is based on the Kolmogorov spectrum of isotropic turbulence, and characterizes the turbulence by a one-dimensional (radial) power spectrum of the form

$$G_1(k) = C_1 \frac{(k\bar{L}_O)^2}{[1 + (k\bar{L}_O)^2]^{11/6}} \tag{15}$$

where C_1 is a constant, k is the wave number, and $L_O = 2\pi\bar{L}_O$ is the so-called "outer scale" of the turbulence (i.e., the size scale at which energy is input to drive the turbulence). It should be stressed that the turbulence spectrum represented in equation (15) is based only on thermally generated atmospheric fluctuations and does not contain contributions at very long wavelengths (small k) due to mechanically generated fluctuations (e.g., by pressure waves, mountains).

A good measure of the astrometric error introduced by atmospheric turbulence can be had by examining the variance in optical path for light from a star as detected at two apertures separated by a distance B. The difference in optical path length, L_2, at two points separated by distance B is

$$L_2(B) = \int_O^h [N(\underset{\sim}{p}, z) - N(\underset{\sim}{p} + \underset{\sim}{B}, z)]\, dz \tag{16}$$

where the integration is over a vertical path extending from the ground to an altitude h in the tropopause, and $N(\underset{\sim}{p},z)$ is the fluctuation in the refractive index of the atmosphere along vector $\underset{\sim}{p}$ as a function of altitude z. The mean value of $L_2(B)$ vanishes as an ensemble average (taken here as a long-term average). However, the variance in L_2 is nonzero in general; this variance is the quantity that sets the limit on the precision of relative astrometry measurements. A rigorous discussion of the determination of the variance in $L_2(B)$, although informative, is mathematically complex and so will not be given here. A rather detailed discussion is given in appendix A. The salient aspects of that detailed discussion are summarized here.

The limiting rms error $\langle\delta\theta\rangle_{\text{rms}}$ in the measurement of angle θ between two stars is related to the variance in optical path of the aperture by

$$\langle\delta\theta\rangle_{\text{rms}} \sim \frac{\langle L_2{}^2(B)\rangle^{1/2}}{B} \tag{17}$$

where B is the telescope aperture (baseline) and $\langle L_2{}^2(B)\rangle$ is the variance. The theoretical analysis given in appendix A, as well as the behavior of measured variance, indicates how path-length variance depends on baseline (fig. 7). The variations of $\log(\langle L_2{}^2\rangle^{1/2}/\lambda)$ (λ is the wavelength of observed light) versus $\log B$ are shown for a range of assumed values of the outer scale L_O of the turbulence and for excellent seeing conditions. For baselines that are small compared to L_O,

$$\frac{\langle L_2{}^2(B)\rangle^{1/2}}{\lambda} \propto B^{5/6} \tag{18}$$

Combining equations (17) and (18) yields

$$\langle\delta\theta\rangle_{\text{rms}} \sim \frac{\lambda}{B^{1/6}} \tag{19}$$

Thus, the limiting error varies linearly with the wavelength of observed light and it varies slowly ($B^{1/6}$) with aperture. Taking

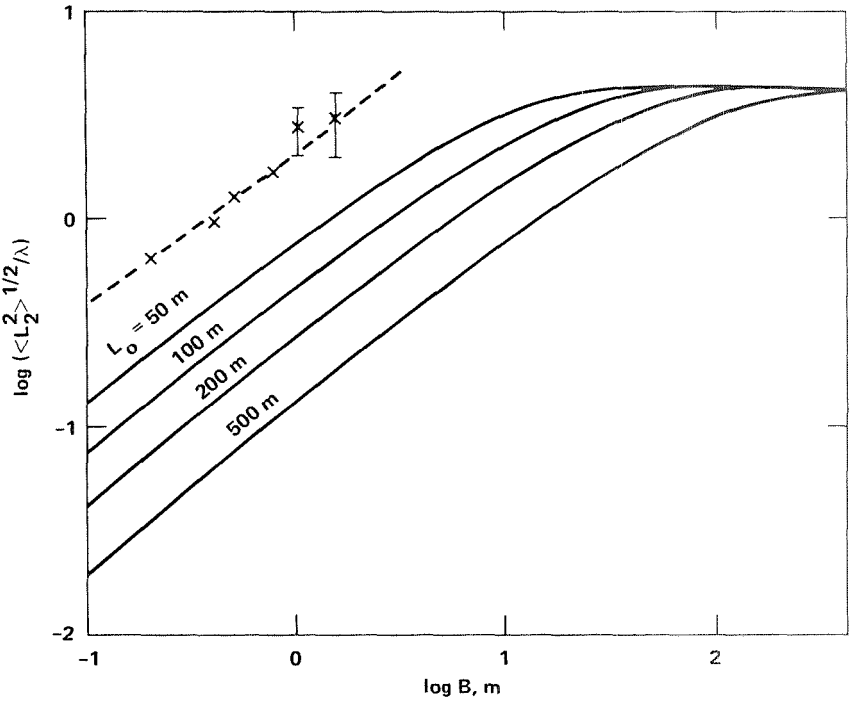


Figure 7.— *A plot of the root-mean-square optical path difference* ($\langle L_2^2 \rangle^{1/2}$) *as a function of interferometer baseline* (B). *Solid curves pertain to differing assumptions concerning the turbulence outer scale* (L_0); *crosses refer to data discussed in appendix A.*

L_O = 50 m, B = 1 m, and $\lambda = 5 \times 10^{-7}$ m gives rise to $\langle \delta\theta \rangle_{rms} \sim 0.08$ arcsec, and $\langle \delta\theta \rangle_{rms} = 0.04$ arcsec for B = 50 m. These values of $\langle \delta\theta \rangle_{rms}$ pertain to the error expected for measurements taken over a time scale that is comparable to the time scale for wind-driven eddies to sweep across the field of view. For 5–10 m/s winds and $B \sim 1$ m, these transit times are <0.2 sec. Additional empirical evidence for this behavior is shown in figure 8, where positional errors for observations on the ~0.8-m Thaw telescope at Allegheny Observatory are shown as a function of observing time. Note that the errors at the short time limit are consistent with the value inferred from the analysis given here. Also evident from figure 8 is that the photographic plate, used as a detector, leads to poorer long-time accuracy than if data were taken in short bursts where the inherently

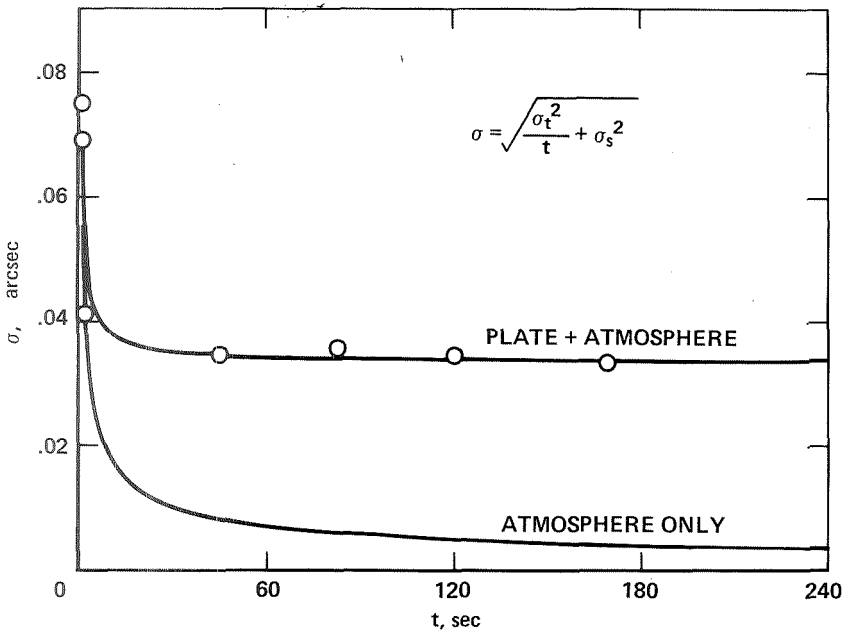


Figure 8.—*Accuracy (σ) of a ground-based astrometric observation using the Allegheny Observatory Thaw telescope shown for various integration times* (t). *Data show that the photographic plate introduces significant errors for long integration times (lower curve indicates errors due only to Earth's atmosphere).*

less accurate but statistically independent atmosphere effects can be used to define a more precise, long-term position. If measurements were taken every τ seconds over a nightly observing time T_{obs}, the nightly rms error $\langle \delta\theta \rangle_{night}$ would be

$$\langle \delta\theta \rangle_{night} \sim \langle \delta\theta \rangle_{rms} \sqrt{\frac{\tau}{T_{obs}}} \tag{20}$$

The theoretical arguments used to derive equation (20) indicate that if $\tau = 0.2$ sec, the expected precision using a 1-m-aperture telescope with a perfect detector is more than an order of magnitude better than for that telescope with a photographic plate as a detector.

The principal conclusion to be drawn here is that Earth's atmosphere places rather fundamental limits on the precision with which

astrometric measurements can be made. The best nightly precision attainable with a 1-m-aperture astrometric telescope and a perfect detector (i.e., one that introduces no astrometric errors) is ~0.0006 arcsec (τ = 0.2 sec and T_{obs} = 1 hr = 3600 sec), and is only ~0.0003 arcsec for a 50-m-baseline telescope. Additional theoretical and experimental studies of the effects of the atmosphere on astrometric precision are clearly needed, but the results given here are unlikely to be much in error (an independent estimate of this limiting precision has been carried out by P. Connes; his findings are consistent with those of Project Orion). Whether very long (~1 km) baselines can be used to increase precision, as suggested by Currie and co-workers, is an open question. The principal uncertainty of long baselines is the effect of mechanically induced atmosphere disturbances.

Comparison between obtainable astrometric accuracy, usually assigned to yearly mean relative positions of stars, and the accuracy required to detect the SPS requires that the nightly precision given above be increased approximately in proportion to the square root of the number of nightly observations. Assuming an average of 36 nights/yr of observing, the limiting accuracy is $\sim 10^{-4}$ arcsec.

DETECTION PROBLEM – HARDWARE ASPECTS

Having discussed the effects of Earth's atmosphere that give rise to errors in relative astrometry observations, we will now discuss errors due to (1) instabilities in the optical components of a telescope and (2) measurement devices. Because the main focus of Project Orion was on astrometric detection of extrasolar planets, the discussion here is limited to astrometric telescopes.

Instability of Telescope Optics

If a refracting telescope is used for astrometry, very small lateral shifts of lenses and changes in their separation may cause appreciable changes in lateral chromatic aberration. The resulting shifts of star images depend on the color of a star. Effects of this kind, amounting

to 0.03 arcsec (ref. 8), led to spurious data concerning the motion of Barnard's star.

The color effects mentioned above are absent in reflecting telescopes. Among the aberrations of reflecting telescopes, coma is most harmful for precise astrometry. Born and Wolf (ref. 9) state, in their classical treatise on optics: "Because of the asymmetrical appearance of an image in the presence of coma, this aberration must always be suppressed in telescopes, as it would make precise positional measurements impracticable." If a nonlinear detector, such as a photographic plate, is used, coma causes shifts of star images, depending on the brightness of a star and its color (since transmittance of Earth's atmosphere is color-dependent) and on the characteristics of the photographic plate. The relative positions of stars can appear to change slowly with time when measured on photographs taken year after year. These apparent changes could be caused by advances in photographic technology, by increasing air pollution, and by climatic changes, particularly those with periods of 1 year and 11 years (e.g., sunspot cycle).

Fortunately, coma of reflecting telescopes can be completely eliminated by using a secondary mirror of suitable shape. However, secondary mirrors are a source of astrometric or positional errors of another kind. Every star in the observed field in the sky illuminates a different area on a secondary mirror. Lateral shifts in position of a secondary mirror and slight changes in the shape of its reflecting surface, occurring over a time scale of years, cause astrometric errors. Such changes in shape may be caused by varying temperature, by relief of stresses in the glass and sagging of the mirror, or as a result of re-aluminizing the surface.

Suppose that a reflecting surface (e.g., a secondary mirror) which is approximately flat is placed between the entrance aperture of a telescope and the image plane, at a distance $(1 - K)F$ from the former and KF from the latter. Here F is the focal length of a mirror or lens placed at the entrance aperture. It is assumed that temporal changes in shape of the reflecting surface remain correlated for points on this surface separated by a distance smaller than s, the correlation distance for that surface. Also assume that the rms relative shift perpendicular to the surface for two points separated by a distance larger than s is $\lambda/100$, where $\lambda \cong 0.5$ μm is the wavelength of visible light. The value $\lambda/100$ is the precision with which surface

deformations can be measured and hence calibrated out, at the present state of the art.

On these assumptions, the tilt of a surface element of diameter s fluctuates with an rms deviation of $\lambda/(100s)$ radians. If an angular separation between the star under study (the program star) and a reference star is ω radians, the centers of areas on the approximately flat reflecting surface which are illuminated by these two stars are separated by $r = (1 - K)F\omega$ and the diameters of these areas are $d = KD$, where D is the diameter of entrance aperture.

Consider two cases: The first case ($d \leqslant s$) occurs when the reflecting surface (secondary mirror) is very close to the image plane. The rms tilt of the reflecting surface situated at distance KF from the image of a star shifts this image as if the star had moved in the sky by $2K\lambda/(100s)$. The change $\delta\omega$ of angular distance ω (in radians) between two stars, that is, the astrometric error caused by tilts of corresponding elements of mirror surface, is

$$\delta\omega \cong 2K\lambda R(\omega)/(100s) \tag{21}$$

where

$$R(\omega) = \begin{cases} r/s & \text{for} \quad r \ll s \\ 1 & \text{for} \quad r \geqslant s \end{cases} \tag{22}$$

In the second case, $d > s$, so that the average number of randomly tilted areas on the secondary mirror illuminated by each star is d^2/s^2. Therefore, the astrometric error is smaller than that in the first case, by a factor of $(d^2/s^2)^{1/2} = KD/s$. This error equals

$$\delta\omega \cong 2\lambda R(\omega)/(100D) \tag{23}$$

If s is smaller than both r and d, the astrometric error caused by surface instability is given (in radians) by

$$\delta\omega \cong 2\lambda/(100D) \tag{24}$$

depending only on the diameter D of the telescope. As an example, for $D = 1.5$ m (Flagstaff astrometric telescope), an error

$\delta\omega \cong \pm 0.0014$ arcsec is produced. This error can be considered as a systematic error slowly changing with time.

Instability of Optical Components of an Imaging Interferometer

If astrometry is based on direct imaging of stars, without interferometry, the position of a star is defined by the position of a centroid of its seeing or diffraction disk. The position of a star becomes independent of the location of this centroid if a Michelson stellar interferometer is used. In the latter case, the position of a star is defined by the location of a white fringe produced by the interferometer, as shown schematically in figure 9. An aberration such as coma affects the white fringe location, but can be calibrated in an interferometer.

A telescope becomes a Michelson stellar interferometer if, instead of its entire entrance aperture, only two portions of this aperture are utilized, each situated a distance $B/2$ from the center of aperture. The distance B is the baseline of the interferometer. Light from each of the small apertures will combine on the image plane to form an image of each star in the field of view. As a consequence of the nature of light, these images will contain interference patterns

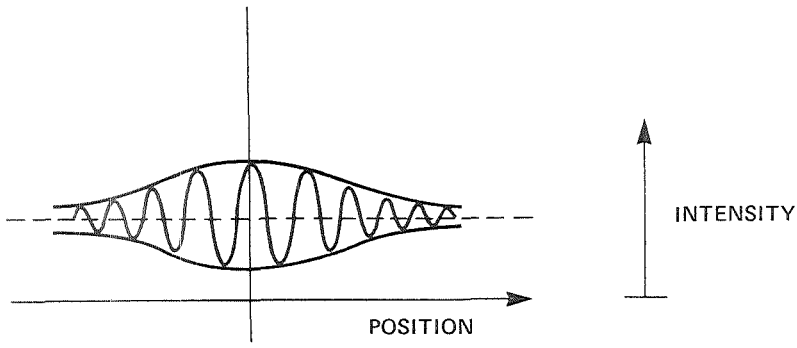


Figure 9.— *Schematic representation of a white-light fringe produced by an interferometer. The location of maximum intensity variation marks the true position. Solid lines forming an envelope about the oscillating line indicate the manner in which the intensity variations decrease with increasing distance from the central fringe.*

comprised of colored, rainbow-like fringes. Only the central bright fringe is white, because it is common to light of all wavelengths. This white fringe is formed by those stellar light rays for which the optical paths through the two arms of the interferometer are exactly equal.

Assume that a secondary mirror is located a distance KF from an image plane. Each star illuminates an area of diameter d on the secondary mirror, and the centers of these areas are separated by a distance r. The rms relative shift perpendicular to the surface is again assumed to be $\lambda/100$ for two points on the mirror separated by a distance larger than s. Let s be smaller than r and d. If the path difference $L_1 - L_2$ between the two arms of the interferometer changes randomly by δL_A for the one star and δL_B for another star, the astrometric error of the angular distance between the two stars is

$$\delta\omega \cong \frac{\left(\delta L_A{}^2 + \delta L_B{}^2\right)^{1/2}}{B} \tag{25}$$

On the assumptions that the rms of δL_1 or δL_2 is $2(s/d)(\lambda/100)$, the astrometric error (in radians) is

$$\delta\omega = \frac{2^{3/2} s\lambda}{100\, dB} \tag{26}$$

For a given baseline B, the error is largest if $s = d$. In this case, for a baseline $B = 50$ m (adopted in ch. 4), the astrometric error caused by slow changes in mirror shape is

$$\delta\omega = \pm 6 \times 10^{-5} \text{ arcsec} \tag{27}$$

If a secondary mirror is cut into two halves, each corresponding to an arm of the interferometer, small relative tilts or shifts of these halves are inconsequential, causing only changes in scale of an image of the star field; the tilts are considered small if star images remain strongly overlapping. Equation (26) applies also to any tracking flat mirrors placed in front of an entrance aperture. In this case, $d >> s$, and therefore the astrometric error $\delta\omega$ is very small.

This discussion, and equations (24) and (26), indicate that an interferometer of long baseline B, as compared to a telescope of diameter D, is less sensitive, by a factor of $2^{-1/2}\ B/D$, to slow changes in the shape of its optical surfaces.

Measuring Devices

Astrometric errors can arise not only in the transmission of light by the optical system of a telescope, but also in the measurement of the parameters of interest. Several components of hardware fall under the rubric of "measuring devices," including the detection system and any postdetection, telescope-independent measuring machines. It will be useful to begin this section with some brief comments on detection systems, followed by comments on techniques for obtaining high precision in measuring angles.

The principal requirements of a detection system for high-precision astrometric observations are: sensitivity, dynamic range, geometric fidelity, and simultaneous detection of the star under study and its reference stars. The level of sensitivity required for planetary detection is not overly demanding in terms of limiting magnitude ($m_v \sim +15$). Photographic plates and photoelectric detectors can easily reach 15th magnitude, given sufficient integration time. However, as remarked in the section on atmospheric limitations to precision, it is desirable to observe in short exposures (<0.1 sec). Because of this short exposure time, the detector must be characterized by both high quantum efficiency and low noise. To obtain astrometric precision at the level required ($<10^{-4}$ arcsec) for a significant search for extrasolar planets, it is necessary that a reference system be used which contains many (10–20) stars (the requirement for many reference stars is discussed in detail in appendix B), and these stars must lie reasonably close (within $\sim 1^\circ$) to the star under study. Allowing for the possibility that there may be reference stars as bright as $m_v \sim +5$ for a 15th-magnitude program star requires a detector with a dynamic range $\sim 10^4$. The characteristic of "geometric fidelity" for a detector means simply that the detector does not introduce any shift in the apparent relative position of the stars (target plus reference).

The tried-and-true detector of astronomy is the photographic plate, which provides for increased integration times over its forerunner, the human eye, which is capable only of integration times of a fraction of a second. However, photographic plates have a very low quantum efficiency (i.e., the fraction of incident photons that is detected), typically less than 1 percent. Photographic plates do not have a large dynamic range (about 100), and are highly nonlinear detectors over that dynamic range. Finally, the geometric fidelity of photographic plates is relatively poor in at least two crucial aspects. First, if a telescope has asymmetrical optical aberrations (e.g., coma), the images of each star in the field of view are distorted differently. As positional measurements from photographic plates are made by locating the centroid of the stellar images as they appear on the plate, this distortion leads to a positional error. The most precise machine for measuring positions on a photographic plate (ref. 10) is capable of locating centroid positions to within ~0.7 μm, which corresponds to an angular error of about 0.007 arcsec for a focal length of 20 m, a factor of about 70 times larger than the effect expected for the SPS. The second aspect of photographic plates that leads to positional error arises from distortion of the plate, both during the interval of time from detection to measurement and during storage subsequent to any remeasurement of a plate. It is very difficult to quantitatively assess the positional error attendant upon this type of geometric distortion. In light of the inherent limitations of the photographic plate (as applied to this problem), it is clear that a more modern, photoelectric detection system must be used.

Advances in technology related to detection of visual light have been significant over the past decade, leading to a plethora of devices such as charge coupled devices (CCD's) and charge injection devices (CID's). No attempt is made here to provide a review of this new technology. Suffice it to say that these new detectors have low noise, high quantum efficiency (>80%), good dynamic range (although quoted dynamic ranges for CCD devices, for example, are not as large as 10^4 at present), and excellent geometric fidelity. The geometric fidelity of these detectors, which is crucial to a search for extrasolar planets, is aided by the fact that positional measurements can be made in situ and essentially in real time. A major drawback to these detectors relative to photographic plates is that they have limited archival ability; photographic plates properly cared for may be

re-examined decades after they have been exposed. However, this archival attribute is judged to be of secondary importance for the task considered by Project Orion.

In the final analysis, any attempt to detect and study extrasolar planetary systems by astrometric observations depends on the precision with which small angles can be measured. It is of little use to circumvent or minimize errors or uncertainties in determining the relative positions of stars due to the combined effects of Earth's atmosphere and telescope optics if the final step in the process introduces significant errors (as is the case with photographic plates and plate-measuring machines). Perhaps the most precise way to measure the relative position of two sources of light is by means of interferometry, where light from the objects under study is modulated by some type of obstructing edge. In concluding this discussion of the factors that led toward the design concept advanced in detail in chapter 4, it is useful to remark on three possible alternative techniques by which the interferometric properties of light might be used to provide high-precision positional data. These techniques involve (1) sequences of gratings, (2) gratings with lenses, and (3) delay lines.

Sequences of gratings– A familiar device that is useful to define angles is a pair of picket fences, that is, two surfaces separated by distance a and obstructing lines in each surface separated by the semiregular distance d. Like a rule inscribed with periodic marks for convenience in length measures, a pair of picket fences is a device that defines periodic directions separated by angle d/a. Such a device is experimentally attractive because of nonsensitivity to misalignments. Nominally, the fence surfaces should be parallel, the laths in different surfaces should be parallel, and the viewing direction should be along the coincident normals to each fence. Let the corresponding tilt, skew, and slant error angles be t, s, and w, respectively. Then, if attention is given to the local value of the distance a between the fences along the viewing direction, the periodic angle becomes $(d/a)\cos t \cdot \cos s \cdot \cos w$. Moreover, when the concept is extended to the use of ruled transmission gratings in parallel light, we note that d can be an extremely precise length and that none of the aberrations apply that trouble angle measures made with the aid of a lens: coma, astigmatism, spherical aberration, or distortion.

Most applications of this technique to date have involved monochromatic light (see refs. 11–17). It seems worthwhile, therefore, to note that broadband light can be used for angle measures *if* it is made periodic in wavelength by passing it through either another grating or a Fabry-Perot etalon. The resulting interferences are comparable to the white-light effects, called Brewster's fringes, that are used to adjust Fabry-Perot etalons to integer length ratios in measures of the meter in terms of light wavelengths.

The grating arrangement depicted in figure 10 leads to light of wavelength λ being diffracted by angle θ, where $d \sin \theta = j\lambda$ and j is an integer. The deflection of light at distance a is $ja\lambda/(d \cos \theta)$. As the wavelength is changed, the light energy that passes grating G2 tends to alternate between orders with change in deflection by distance d. Therefore, the light needs to be periodic with interval $\Delta\lambda = d^2 \cos \theta/(|j|a)$. Such a relation arises in a Fabry-Perot etalon of length L, where L is determined by

$$\Delta\lambda = \frac{\lambda^2}{2L}$$

giving rise to

$$L = \frac{a}{2|j|} \frac{\lambda^2}{d^2} \sec \theta$$

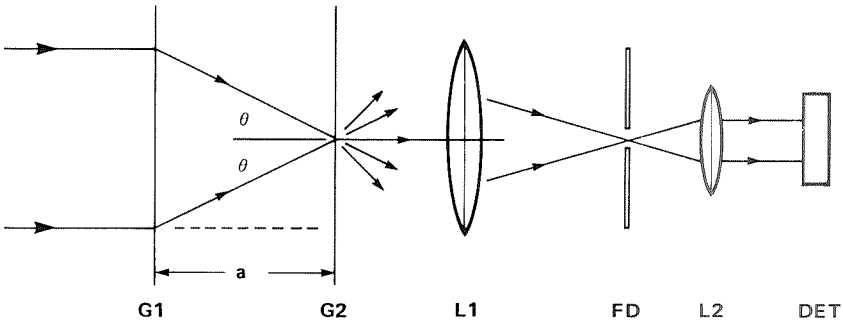


Figure 10.– *Schematic representation of a detector system employing a sequence of gratings to define accurate position measurements. Light incident on gratings G1 and G2 is focused by lens L1 on a field-defining aperture (FD) and refocused by lens L2 on a detector (see text for further discussion).*

The periodic energy distribution along grating G2 that is thus selected constitutes a self-imaging of grating G1, known as the Talbot effect (refs. 18 and 19), that persists even if the incident light is not collimated (ref. 20) because the reconstruction is holographic. Note that $\theta \cong \lambda/d$ for first-order diffraction and that, if θ is small, then L can be so small that the Fabry-Perot can be constructed by thin-film techniques in many cases of interest.

A new technology relevant to these remarks is the advent of holographic phase gratings whose surface amplitude is chosen so that the energy in the undiffracted zeroth-order beam is vanishingly small for a small range of visible wavelengths. The energy from one diffraction is mostly in orders of ± 1. After light passes through two such gratings, the energy alternates mostly between orders zero and ± 2. The latter orders are easy to eliminate and the zeroth order can be fully modulated. Furthermore, the gratings can be physically large ($\sim$0.2 m).

Gratings used with lenses– Ronchi (ref. 21) has reviewed the theoretical development and use of a single small grating near the focus of a lens, as depicted in figure 11. The plane grating at distance q from a focal plane of the lens in the light from a point source causes light to pass through points P_1 and P_{-1} in addition to P_0 in the focal plane. At an observation screen at distance r beyond the focal plane, there are intensity oscillations (fringes) in the overlap region BE. The sources P_0 and P_1 are separated by approximately $y = q(\lambda/d)$. Such sources are coherent since they are derived from the same source. Therefore, the light fields reinforce at separations z on the observing screen, where $z = r(\lambda/y) = rd/q$. Note that if $r = q$, the separation distance z equals the grating spacing, independent of light wavelength λ. The display seems to be a rectilinear projection of the grating onto the screen. But a wave-optical formulation is essential, as shown by a fringe spacing of $z/2$ in the region CD illuminated by P_1 and P_{-1}, because the interfering sources are separated by $2y$. More generally, many spacings of the form $z, z/2, z/3$, etc., may be present if r/q is large and many orders interfere. Notably, the fringe spacing for all orders of interference becomes infinite as the grating moves into the focal plane. If the grating is behind the focal plane so that q in effect changes sign, sources P_1 and P_{-1} become virtual.

Malacara and Cornejo (ref. 20) showed that, to obtain constructive interference for all orders on the observing screen, distance q

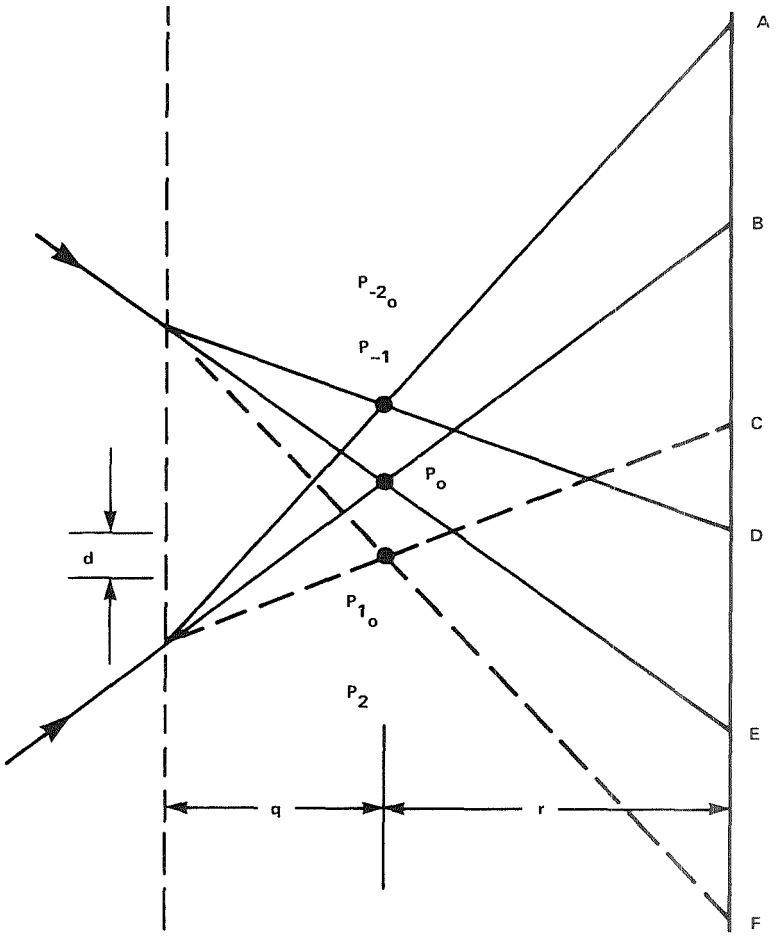


Figure 11.– *Schematic representation of a detector system employing a self-imaging grating system (see text for further discussion).*

should be an integer multiple of the Rayleigh distance $2d^2/\lambda = d/(\lambda/2d)$. Such a length emerges in Talbot self-imaging when the diffraction involves mainly zeroth-order transmission and weak ±1 orders.

The two methods discussed above provide alternative means by which light from a source can be manipulated so as to produce a well-defined interference pattern. The next step involves manipulation of such a fringe pattern to allow measurement of the separation between two light sources.

A plane grating near the focal plane of a telescope can be used as a modulator of star images to measure distances in the focal plane. If two star images are separated by a distance $(i + g)d$, where i is an integer and $0 \leqslant g < 1$, then a grating modulator produces two waveforms that are similar, but one of which is delayed in time by the fraction g of a modulation cycle (see fig. 12). The grating motion can be linear or vibratory, depending on whether the grating perfection is locally accurate enough for the measurement purpose. Linear motion permits averaging over many grooves. The second coordinate can be obtained by rotating the grating by a known angle or using a grating with grooves in two directions. Note that when a position accuracy of ω radians is desired, the rotation angle between the modulation directions must be known to ω radians. If a mechanical rotation were planned, the observer could provide for repeatability of the angles by arranging to look through a two-direction grating at the rotated grating. The moiré fringes of nearly aligned grating grooves permit highly accurate rotation settings.

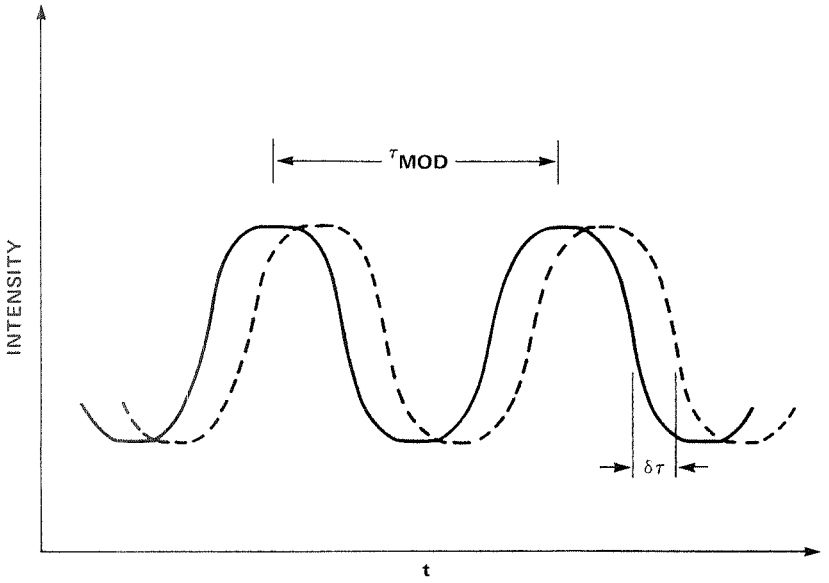


Figure 12.– *Schematic representation of the intensity variation for two phototubes exposed to light from two stars. Modulation is achieved by passing an alternating opaque/transparent Ronchi ruling over the phototubes.*

Two principal approaches to the grating modulator emerge. A grating of good local perfection can be placed in the telescope focal plane, or two gratings can be used to *synthesize* a grating of good local perfection in the focal plane. The former might be a Ronchi ruling or holographic amplitude grating. If the grating spacing is less than the apparent star image diameter, there must be several detectors within the pupil. The synthesized grating could be obtained by placing two gratings symmetrically about the focal plane. Then the following pupil contains many fringes because of interferences between sources P_O, P_2, and P_{-2_O} (shown in fig. 11). If the many detectors, which are thus required in the pupil, are objectionable, one can form an image of these sources, block all but P_O, and put one or more detectors in the pupil that follows. Clearly, the designer of a grating modulator who uses the principles of Abbe imaging will have much flexibility in that design.

Delay lines– Miller (ref. 22) suggested in a context of small-angle measures that the angles might be measured with high precision in the geometry of figure 13. If mirrors of fixed separation are used

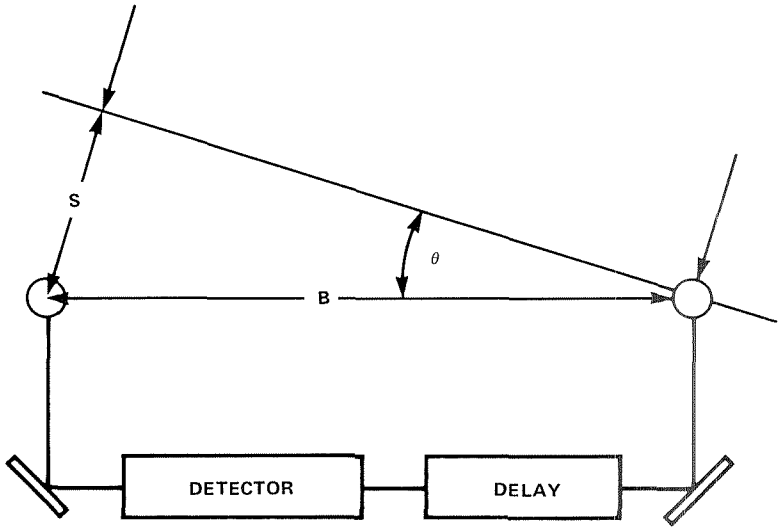


Figure 13.– *Schematic representation of a stellar interferometer. Light collected by two apertures separated by a distance* B *has a time delay* S/c, *where* S = B *sin* θ. *This time delay due to longer path must be compensated by a delay in the detector.*

to collect light, then angles can be measured by introducing changes in a delay line. As usual, angles are inferred from ratios of lengths in a geometric construction. In recent years, the technology of infrared spectroscopy has developed precise and reliable delay lines.

It is important to realize that the geometry in figure 13 is completely equivalent to imaging with a lens. The delays appropriate to a lens arise from suitably placed reflecting or refracting bodies and free-space propagation to the image point corresponding to each object point. But when the lens is uncomfortably large so that only small elements of the lens can be furnished, the necessary delays can be furnished by various means. The path lengths can be monitored by laser interferometers so that high accuracy is attained. The use of delay lines for high-precision relative astrometric angle measurement is currently under consideration by Currie and co-workers.

The use of any optical system to precisely measure angles between light sources involves the interference properties of light; the number of such systems that are adequately accurate is large. The essential factors controlling interferometer design are nicely stated by Ronchi (ref. 21): "Finally, fringes given by the grating interferometer bring one to the same conclusions that are reached with the use of any interferometer whatsoever: the difference between one interferometer and another in the testing of optical systems is reduced to a question of *practicality, rapidity,* and *economy*."

SUMMARY

The fundamental aspects of the planetary detection problem, namely, the physical basis of the observable phenomena and the effects of Earth's atmosphere and telescope hardware on observations, have been discussed. The primary emphasis of that discussion was to focus on those factors which set the constraints that influence the formulation of design concepts. Primary consideration was given to those aspects of the detection problem that might affect the design of a ground-based astrometric telescope. The major conclusions, as they relate to defining the foundation of a design concept for such a telescope, are summarized below.

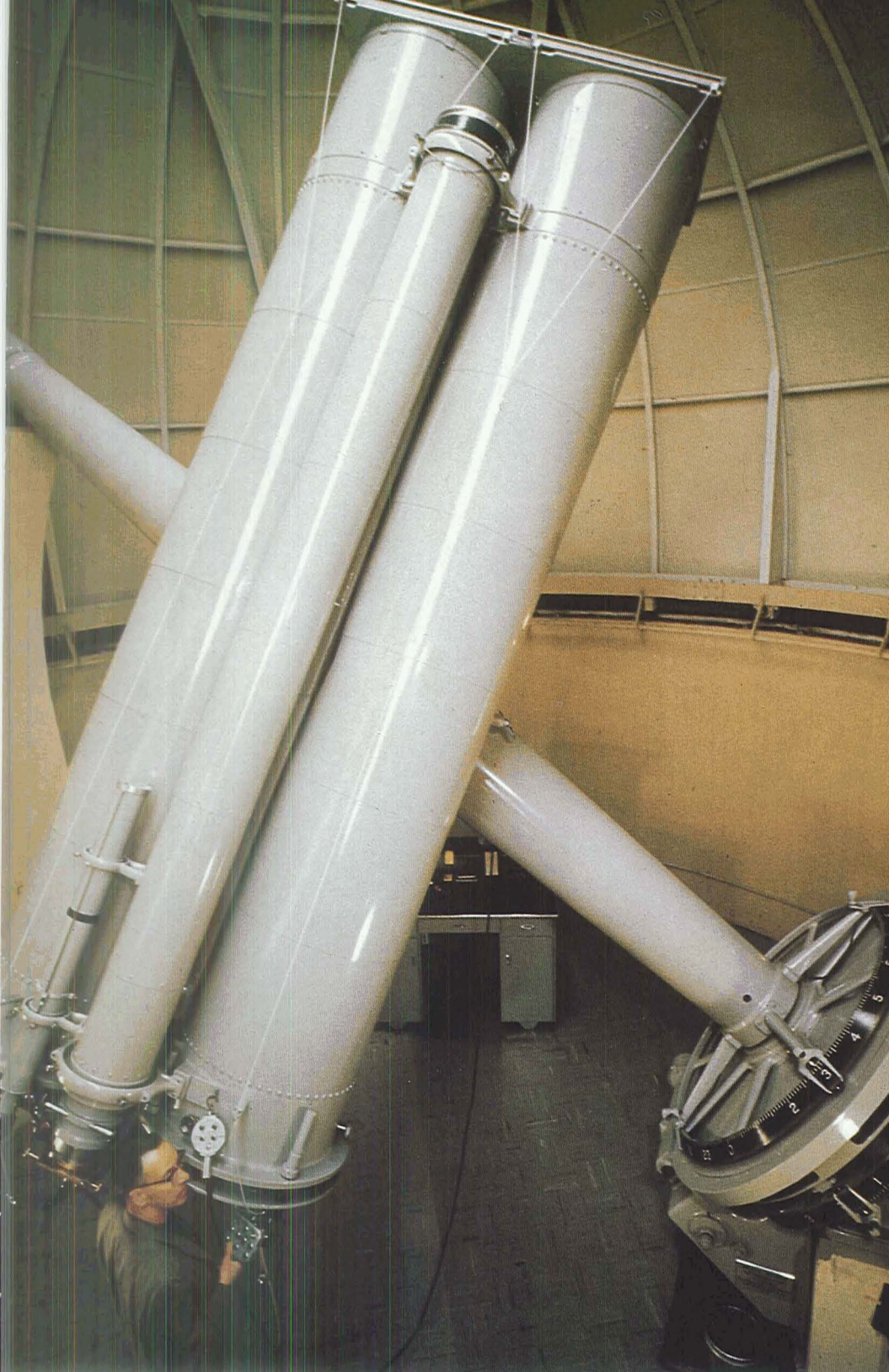
1. Astrophysical aspects
   - Expected angular displacement is small, beyond the measurement capability of the best current astrometric telescope
   - Anticipated observing time comparable to planetary orbital periods (long-term stability required)
2. Terrestrial aspects
   - Atmospheric thermal turbulence limits ground-based astrometry to a precision of $\geqslant 10^{-4}$ arcsec
   - Highest accuracy obtained from taking many independent, short ($\tau < 0.2$ sec) observations/night rather than continuously integrating
   - Important length scale for thermal turbulence is about 50–100 m; therefore an aperture or baseline of at least this size is desirable
   - Highest accuracy is obtained from measurements that minimize refractive effects of the atmosphere (i.e., measurement in azimuth)
3. Hardware aspects
   - Optics
     - Reflecting optics preferred over refracting optics because of absence of chromatic effects
     - Reflecting optics subject to coma; however, coma can be eliminated by suitably configured secondary mirrors
     - Changes in position or figure of secondary optical elements give rise to astrometric errors; sensitivity of telescope to these errors varies inversely with aperture (or interferometer baseline)
   - Measuring devices
     - Requirements of low noise, high sensitivity, dynamic range, and geometric fidelity rule out photographic plates as detectors
     - Precise measurements of angle between two stars best obtained from measurements of the relative linear positions of white fringes produced by interfering light from each star

The findings listed here indicate that a promising design concept for a high-precision, ground-based astrometric telescope is one which

has (1) reflecting optics, (2) large aperture (to minimize sensitivity to astrometric errors arising from secondary mirrors and to minimize atmospheric effects from turbulence), and (3) "electric" detection of interference fringes from stars. One very important additional characteristic is that the telescope simultaneously image many stars (10 to 20). (This characteristic has not been discussed in detail in this chapter, but it is discussed later.) Briefly, a large number of stars is essential to obtaining enough conditional equations to determine parameters that permit modeling of both short- and long-term drifts in the data. Simultaneous imaging is essential to minimize variance of those parameters to the extent that they do not limit the attainable accuracy of the system.

A design concept which embodies the above elements is that of a long-baseline interferometer that simultaneously "images" the white light fringes of many stars. Such a concept is the one conceived and developed during Project Orion. Details of the Orion imaging stellar interferometer are given in chapter 4.

20-inch Astrograph
Photograph courtesy of Lick Observatory, University of California, Mt. Hamilton, California.



3. DIRECT DETECTION SYSTEMS

The information content intrinsic to indirect (astrometric) detection techniques and to direct (at visual or infrared wavelengths) detection techniques was discussed in chapter 2. An important aspect of that discussion is that indirect and direct detection techniques are complementary in terms of information content. In addition, direct and indirect techniques are also complementary with regard to the types of planetary systems that the techniques are best suited to discover. Direct techniques are best suited for detecting bright (or hot) planets, which would be found around more massive bright stars. Indirect techniques are best suited for detecting planets that revolve around low-mass stars. In view of the possible role direct detection techniques could, and should, play in a comprehensive program to search for extrasolar planetary systems, a small effort was made during Project Orion to consider direct detection systems. The results of that effort are presented in this chapter.

DIRECT DETECTION AT VISUAL WAVELENGTHS

It was pointed out in chapter 2 that direct detection almost certainly must be done from space. No consideration was given during Project Orion to design aspects of potential spacecraft that might house a telescope that could undertake a search for visual light reflected by extrasolar planets. Rather, consideration was confined to defining a system concept that might serve as a baseline for more detailed future studies.

The resolving power Θ of a telescope, as a function of aperture D and wavelength λ of the observed light, is given by

$$\Theta(\lambda,D) = 1.22\,\frac{\lambda}{D} \tag{28}$$

This relation is generally attributed to Rayleigh and hence is known as the Rayleigh criterion. Equation (28) shows that a very modest telescope in space, $D \cong 0.25$ m, would have an angular resolution at

$\lambda = 5 \times 10^{-7}$ m of 0.5 arcsec, adequate to resolve the planet-star pair in the SPS. However, the Rayleigh criterion is applicable only to resolving two sources of *equal* intensity. It does not adequately describe the relation $\Theta(\lambda, D)$ for cases when the intensity ratio, $\mathcal{R}$, between two sources is very small, as is the case in the SPS.

The transmission of light in a normal clear-aperture telescope is constant over the aperture and goes abruptly to zero at the edge of the aperture. This produces a classical intensity diffraction pattern of the form $(2J_1(R)/R)^2$, where J_1 is the first-order Bessel function and R is the distance from the optical axis. If the light is monochromatic, all of the higher order maxima would be present; however, in most practical situations a range of wavelengths is involved, causing all but the first 10 maxima to be smeared out. The maxima outside the central peak are often referred to as the rings of the diffraction pattern. The intensity produced by a clear aperture at the center of the image, $I(0)$, is given by

$$I(0) = \left\{\pi r_O^2\right\}^2 \tag{29}$$

where r_0 is the radius of the aperture. The intensity $I(\theta)$ at a point off the optical axis, by an angle θ, is given by

$$I(\theta) = \left[\frac{\lambda r_0}{\theta} \cdot J_1\left(\frac{2\pi r_0}{\lambda}\,\theta\right)\right]^2 \tag{30}$$

where θ is in radians. The quantity $I(\theta)/I(0)$ for monochromatic light is shown in figure 14 as a function of Airy radii. For values of θ such that $I(\theta) \ll I(0)$, an asymptotic expansion of $J_1^2(X)$ may be used to give $I(0)/\bar{I}(\theta)$:

$$\frac{\bar{I}(0)}{\bar{I}(\theta)} \sim 3 \times 10^{-15}\left[\frac{D_0\theta}{\lambda}\right]^3 \tag{31}$$

where $\bar{I}(\theta)$ is the mean intensity at θ and $D_0 = 2r_0$. Equation (31) may be inverted to solve for the product $D_0\theta$ (in marcsec (for λ, in m)). As $I(0)/\bar{I}(\theta) \sim \mathcal{R}^{-1}$, the aperture size D_0 is given by

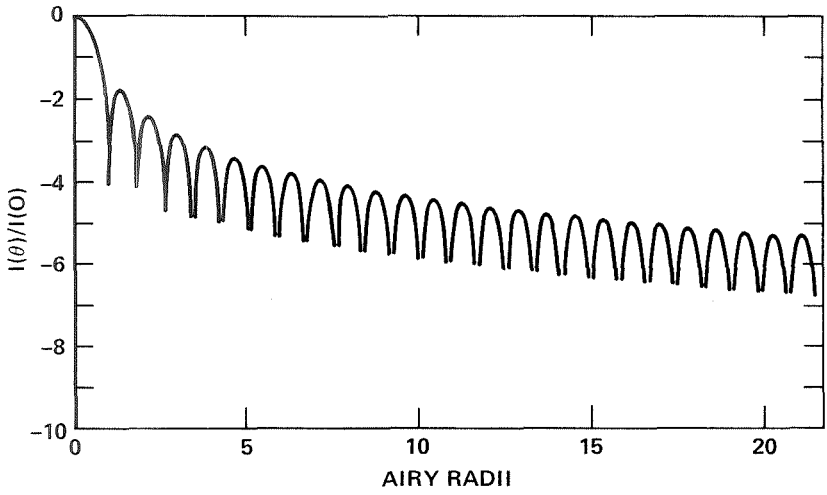


Figure 14.– *Schematic representation of the variation in light intensity, I(θ), in a clear-aperture telescope as a function of distance from the optical axis (θ = 0). Off-axis distance is expressed in terms of Airy radii; the light is assumed to be monochromatic.*

$$D_0 \sim 7.11 \times 10^4 \frac{\lambda}{\mathcal{R}^{1/3}\theta} \text{ m} \tag{32}$$

To resolve the SPS ($\mathcal{R} = 2 \times 10^{-9}$, $\theta = 0.5$) with $\lambda = 5 \times 10^{-7}$ m, a telescope must have an aperture of 56 m! Obviously, a clear-aperture telescope, even though located in space, is not the instrument to use in a search for extrasolar planets. Other approaches must be used to provide the necessary $\mathcal{R}$-value at $\theta \lesssim 0.5$ arcsec.

One approach is to reduce or eliminate the ring maxima that occur outside the central, bright region of the diffraction pattern. A device that accomplishes this is called an *apodizer*. The first thorough examination of the problem of resolution in the limit of very small $\mathcal{R}$-value was by Jacquinot (ref. 23), who first coined the phrase "apodization." (Jacquinot referred to the rings in a classic diffraction pattern as feet, and the process of minimizing or removing the light contained in the rings ("feet") as apodizing, or removing the feet.) As the rings arise due to the abrupt termination in the transmission at the edge of an aperture, the concept of apodization centers on

techniques that produce a more gradual transition to zero transmission at the edge. Two types of apodizer seem most feasible. One is a diaphragm that alters the shape of the pupil; the second is a filter whose transmission varies with distance from the optical axis. Either type must be located at a pupil plane (i.e., any real plane conjugate to the telescope aperture) in order that light rays from every object point will be distributed over the apodization device in exactly the same way as at the center of the distribution.

A second approach toward a direct imaging system involves reducing the bright, central Airy disk. The simplest means, conceptually, to accomplish this is to place an obstruction at the focal plane, or at some plane within the optics which is conjugate to the focal plane. There are, however, objections to this approach. The presence of such an obstruction makes both alignment and pointing of a telescope difficult. When a program or target star is lined up perfectly, it can no longer be seen. Also, the central obscuration would scatter light into the region where a potential planet's image would be located. This scattering can be reduced somewhat by apodization, but it raises the noise level with which apodization must contend.

Another method of reducing the effect of stellar radiation without affecting the planetary light is by means of an interferometer. Planetary visible light is reflected stellar light. Light from the planet has traveled a greater distance to arrive at the observer than has light coming directly from the star. The important point is that the extra length in optical path is large compared to a coherence length,[2] so that stellar light and reflected planetary light will not interfere. It is possible to place an interferometer in the optical system of a telescope so that it would destructively interfere the

[2]The idea of coherence length can be explained in terms of the Michelson interferometer in which a beam of light is divided into two by a beamsplitter. Each beam is directed down an arm of the interferometer to a plane mirror, then it is reflected back to the beamsplitter where the two beams are recombined. There interference fringes are formed by introducing a slight tilt in one of the mirrors. When the lengths of the two arms are equal, the contrast of the fringes is greatest. If one of the arms is lengthened, fringe contrast drops off until at some point they disappear completely. This displacement, the *coherence length*, is a property of the light source. A well-filtered thermal source barely exhibits interference and therefore has a coherence length near zero. A mercury discharge lamp will exhibit a coherence length of several centimeters while a helium neon laser has produced interference over several hundred meters.

stellar light with itself, but not interfere with the planetary light. This operation should be carried out before the light reaches the image plane of the telescope.

A telescope containing these optical elements has special requirements. If the telescope is to be carried into orbit on a shuttle, it must be light and compact. Moreover, it must have a real pupil plane, conjugate to the dominant aperture, where the light-reducing elements may be located. These requirements strongly indicate that such a telescope would be a variant of a folded Gregorian. Aspects of one such system concept are discussed later in this chapter.

Apodization

There are many possible transmission functions that can, in principle, be used to apodize a telescope. One family of functions, which has been studied by Oliver (unpublished data, 1976) in the context of detecting extrasolar planetary systems, is the so-called Sonine family. These functions lead to transmission $T(r)$ over an aperture of the form

$$\left.\begin{aligned} T_\mu(r) &= \left(1 - \frac{r^2}{a^2}\right)^\mu \qquad r \leqslant a \\ T_\mu(r) &= 0 \qquad\qquad\quad r > a \end{aligned}\right\} \tag{33}$$

where a is the aperture size and μ is a positive semidefinite ($\geqslant 0$) index. The aperture form characterized by equation (33) is rotationally symmetric about the optical axis. The tests performed during Project Orion were designed to test the apodization concept by means of aperture masks. The form for such masks is discussed in appendix C.

A variation on the concept of a tapered transmission function is that of an opaque mask at the aperture. Although such a mask does not generally give rise to a circularly symmetric tapered transmission, it does lead to the transmission being reduced smoothly in an average sense in certain preferred directions. The "directionality" of a given mask depends on the geometry of that mask.

A well-known application of the mask technique was the study of the Sirius A-Sirius B system by Lindenblad (ref. 24) and van

Albada (ref. 25). Both workers used a regular hexagonal mask on a clear circular aperture. The diffraction pattern produced by the brighter member (Sirius A) of the pair is like a snowflake (cf. fig. 18). A clear aperture produces a white-light diffraction pattern where the average intensity decreases as the inverse third power of the distance from the optical axis (cf. eq. (31)). The hexagonal mask gave rise to an intensity distribution that decreases more slowly (as the inverse second power) in the spikes of the pattern, but which decreases more rapidly (as the inverse fourth power) in the interspike regions of the pattern. By rotating the principal axes of the hexagonal mask, Lindenblad was able to place one of the relatively dark, interspike regions in register with the location of the image of Sirius B, thereby recording the first direct detection of the B-component. It should be noted that the angular separation of Sirius A from Sirius B is about 10 arcsec and the brightness ratio $\mathcal{R}$ is $\cong 10^{-4}$.

Detecting Sirius B is many orders of magnitude (6–7) easier than detecting the SPS. Further, the hexagonal mask or any aperture edge obstruction is unlikely to give rise to a diffraction pattern with the required spatial-intensity resolution. Masks that vary smoothly are more likely to provide the required diffraction pattern. In this connection, masks predicated on Sonine functions are potential candidates for apodizing to achieve an acceptable diffraction pattern. Although not referred to as "Sonine masks," such masks were studied theoretically by Tuvikene (ref. 26), and further analyzed and utilized by van Albada (ref. 25). In view of the potential of such mask geometries for apodization, an empirical investigation was conducted during Project Orion to ascertain the performance of some rather simple masks.

Apodizing Test Model

The time available for Project Orion precluded an in-depth experimental investigation of the problem of apodization. Consequently, it was decided to utilize the facilities available at Ames Research Center, specifically a 150-m-long dark tunnel, to test a limited number of apodizing masks on a 0.076-m-aperture telescope loaned to Project Orion by Stanford University. Figure 15 is a block diagram of the experimental system. At one end of the 150-m tunnel

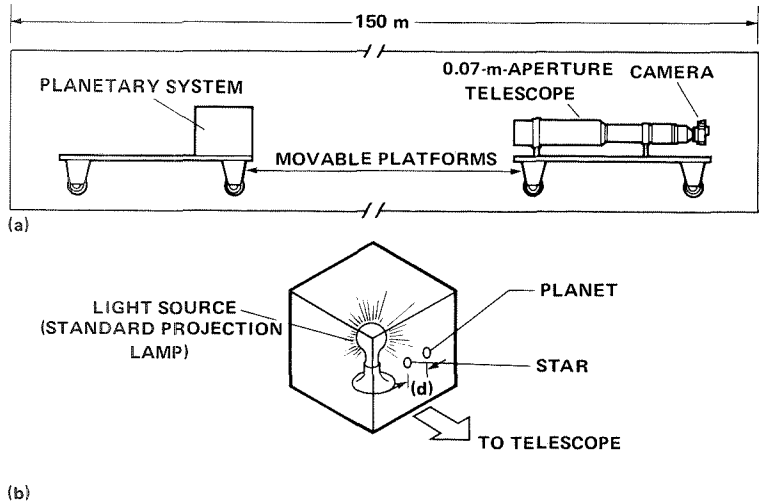


Figure 15.– *Schematic representation of the model planetary system used in the apodization tests. (a) Relative placement of the model planetary system and the telescope/camera system. (b) Light source and two pinholes used to mimic a star-planet system. Various neutral-density filters were placed over the planet pinhole to reduce the intensity of the emergent light.*

was a model planetary system, comprised of a projection lamp in a box with two pinholes (fig. 16). The pinholes were identical in size, and the intensity of light transmitted through one of the pinholes (the "planet") was controlled by neutral density filters. The telescope (fig. 17) was mounted 150 m away from the model planetary system. Masks made from heavy construction paper were placed over the objective lens of the telescope. A camera was mounted at the eyepiece and time exposures were taken using Kodak Plus-X pan film.

The first test involved a model of the Sirius A-Sirius B system. As remarked above, this system is not a severe test of apodization, but does afford a test of the modeling scheme. The results of the Sirius test are shown in figure 18. The model Sirius B is clearly visible at ten o'clock in a null in the diffraction pattern. The mask used for this test was a regular hexagonal mask of the type used by Lindenblad. The effective "resolution" of the model telescope system is far more than adequate to resolve the model binary system, but the characteristic six-spike diffraction pattern is clearly evident.

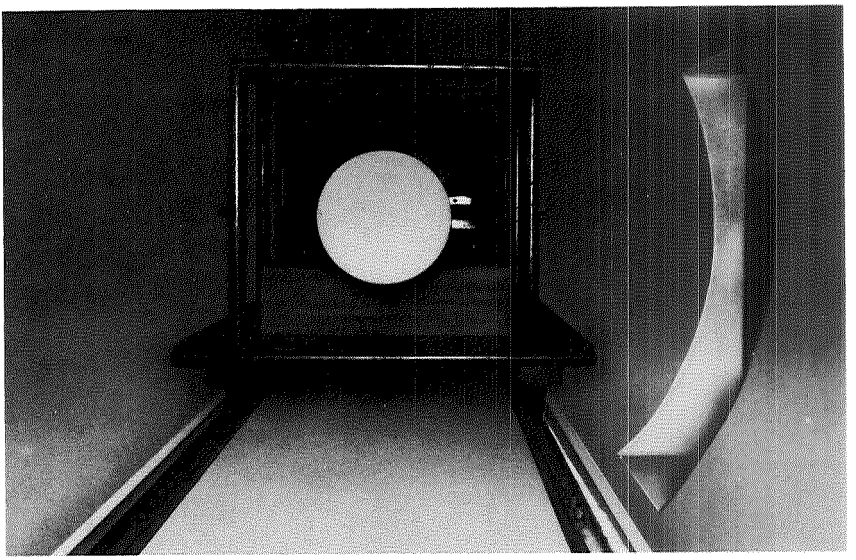


Figure 16.– *Photograph of the "planetary system" in the test tunnel.*

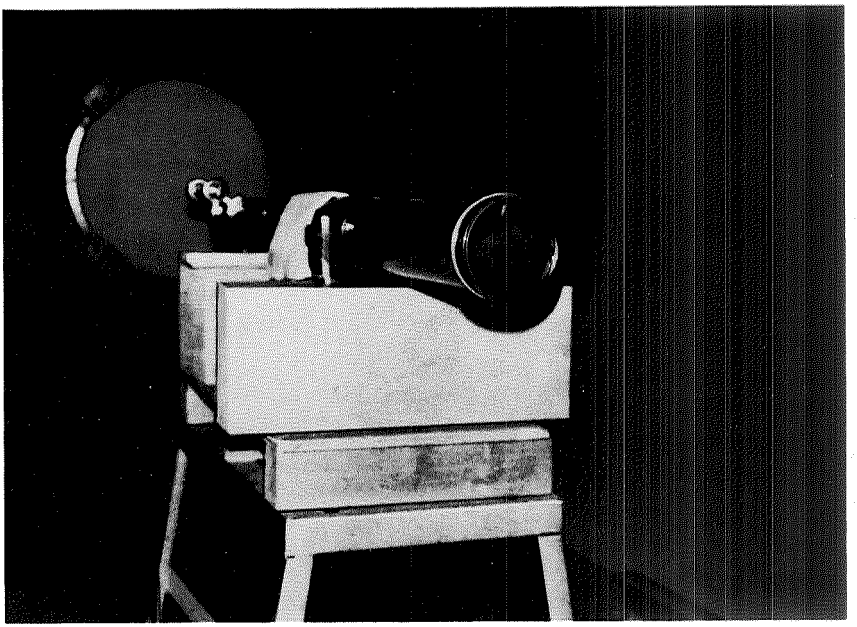


Figure 17.– *Photograph of the telescope used in the apodizing tests.*

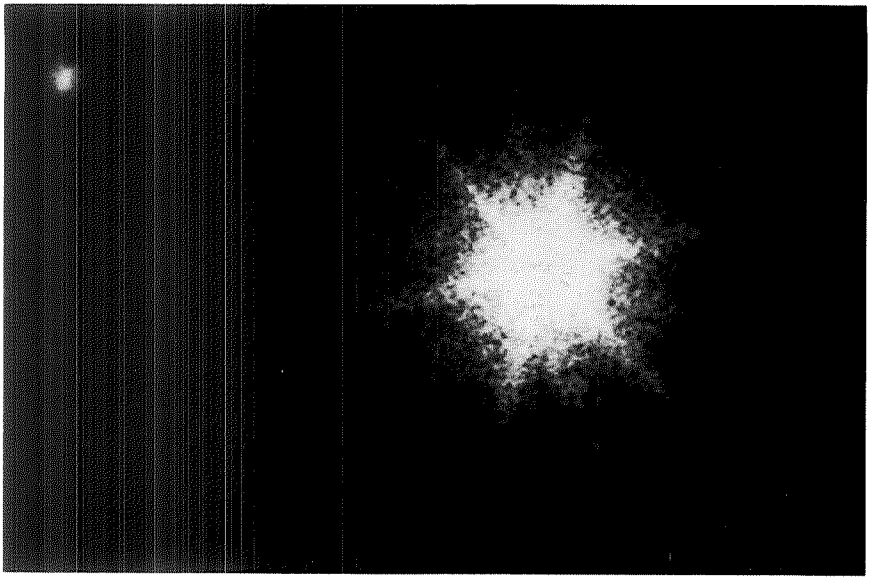


Figure 18.– *Model of the Sirius A-Sirius B system using a hexagonal apodizing mask. Model Sirius B is ~10^4 times dimmer than model Sirius A (see text for further discussion).*

Having demonstrated that the test system yielded qualitatively correct behavior, tests were then run on the scale model of the SPS. As the telescope concept developed during Project Orion for direct visual detection of extrasolar planets has a 1.5-m aperture (see discussion below), the tests were run to simulate the resolution of such a telescope in space. The 0.5-arcsec angular separation that characterizes the SPS corresponds to a linear separation of 3.6×10^{-4} m at a distance of 150 m. The resolution of a diffraction-limited telescope varies inversely as the aperture of the telescope (cf. eq. (28)); thus a 1.5-m aperture would have about 20 times the resolution of a 0.076-in. aperture, such as used for these tests. Including this factor in the scaling of the model SPS leads to a linear separation of 0.007 m at a distance of 150 m.

Tests were run with various aperture configurations, namely, a clear aperture, an aperture with a Sonine mask ($\mu = 7$), and an aperture with the complement to the $\mu = 7$ Sonine mask. Figures 19 through 23 are photographs taken of the model planetary system

using various aperture configurations (fig. 19, clear aperture; fig. 20, Sonine mask; figs. 21–23, Sonine complement) and various $\mathcal{R}$-values (10^{-3}, 10^{-3}, 10^{-4}, 10^{-4}, and 10^{-6}, respectively). The general result is that the Sonine complement aperture mask gave better results than the Sonine mask which, in turn, gave better results than the clear aperture. The Sonine aperture was unable to resolve the model planet at an $\mathcal{R}$-value of 10^{-4}. The results with the Sonine complement mask were as expected in that the distribution of light on the image plane was complementary to the distribution from the Sonine aperture. However, there was a quantitative distinction in the sense that relatively more light was concentrated in the bright areas and relatively less light fell in the null regions of the diffraction pattern. As a consequence, proper orientation of the complementary occulting mask permitted visual detection of the planetary companion at an $\mathcal{R}$-value of 10^{-6}. Unfortunately, the object is sufficiently faint that reproduction in this report washes out the planet; it is, however, clearly visible on the film negative.

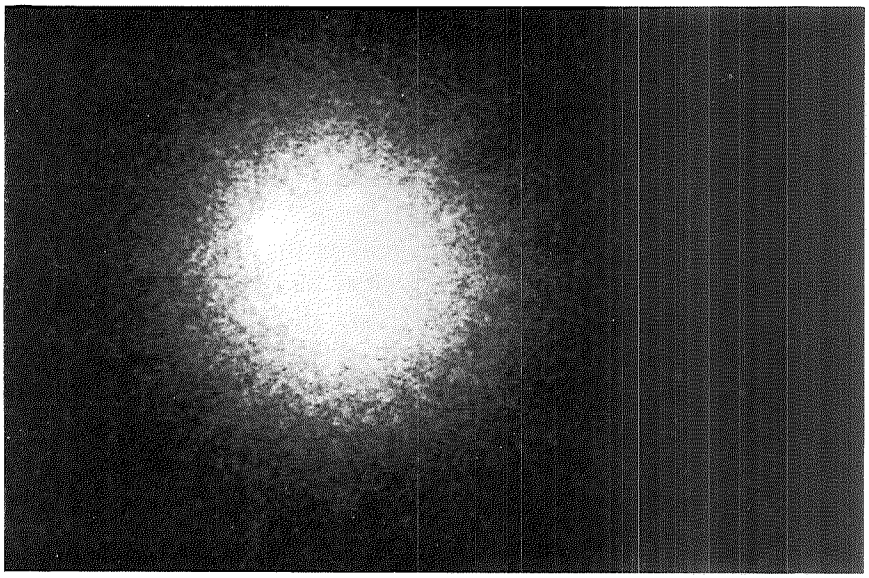


Figure 19.— *Photograph of a model planetary system using a clear aperture (no apodizer);* $\mathcal{R} = 10^{-3}$ *for this system.*

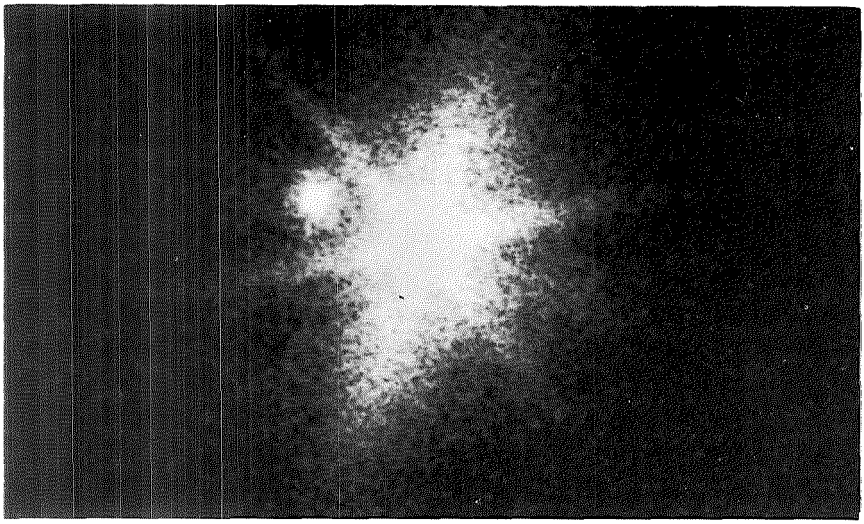


Figure 20.– *Photograph of the same system as shown in figure 19, but with a* $\mu = 7$ *Sonine apodizing mask. Note the enhanced visibility of the "companion" in this photograph.*

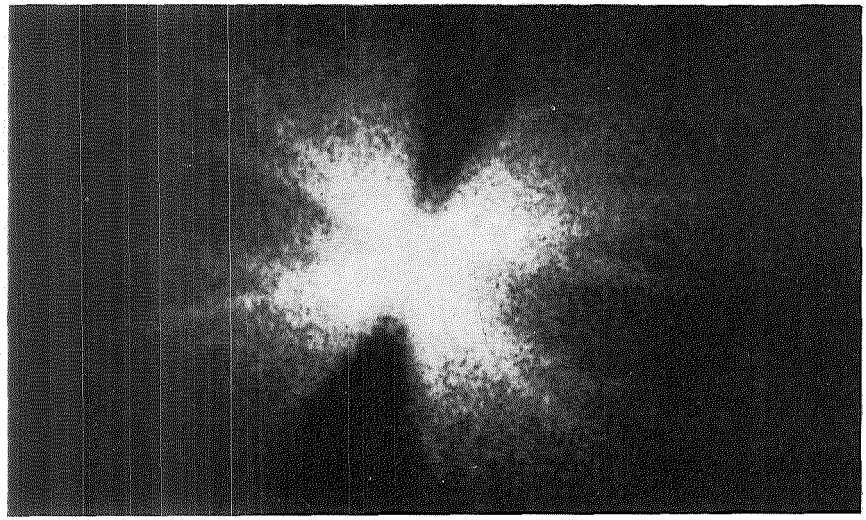


Figure 21.– *Photograph of a model planetary system* ($\mathcal{R} = 10^{-4}$) *using a mask that is the complement of the mask used in figure 20 (see text for discussion). One of the bright spikes obscures the companion.*

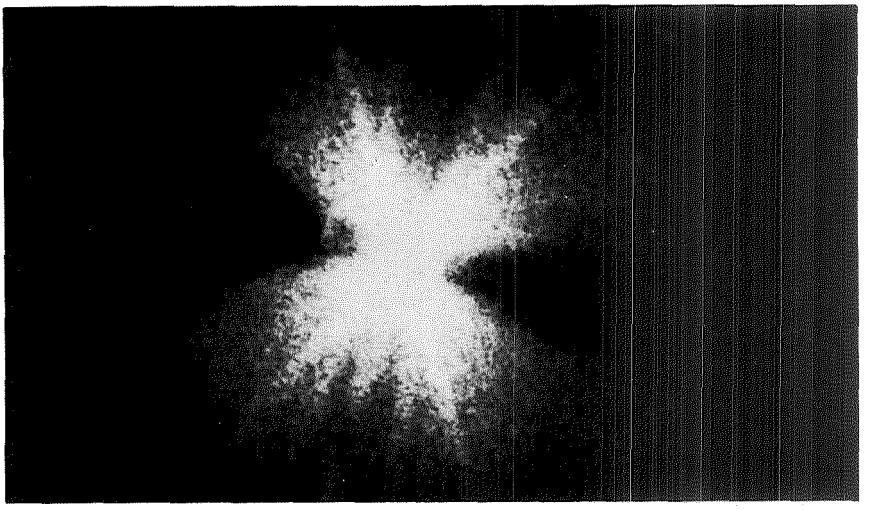


Figure 22.– *Photograph of a model planetary system ($\mathcal{R} = 10^{-4}$) using a mask that is the complement of the mask used in figure 20 (see text for discussion). The mask has been rotated to align a dark interspike region with the location of the companion.*

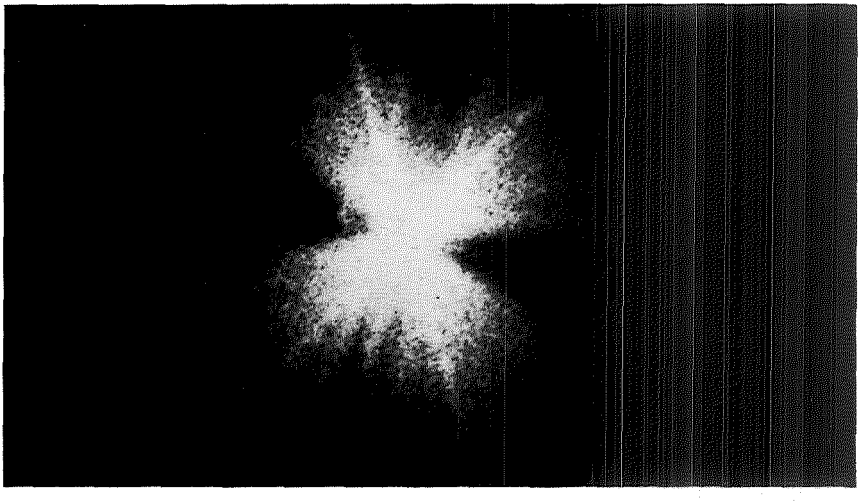


Figure 23.– *Photograph of a model planetary system ($\mathcal{R} = 10^{-6}$) using a mask that is the complement of the mask used in figure 20 (see text for discussion). The mask has been rotated to align a dark interspike region with the location of the companion.*

These tests are not, nor were they intended to be, a definitive and exhaustive exploration of the apodizing problem. They do, however, clearly indicate that relatively unsophisticated masks provide significant off-axis light suppression. It is not unreasonable to suspect that more sophisticated masks, optics, and detectors could reach $\mathcal{R}$-values of 10^{-7} to 10^{-8}. Although these $\mathcal{R}$-values are higher than that of the SPS, it seems possible to obtain the additional two orders of magnitude by, as mentioned previously, combining an apodizer with a system that can reduce the intensity of the stellar Airy disk. (It is tempting to carry the physiological analogy of the diffraction structure to an extreme by associating the Airy disk with the "head" and denote a device that removes the head as an acephalizer; that temptation will, however, be overcome.)

Star Cancellation

As discussed above, a combination of an apodizer and a device to cancel or minimize the light from the star that might reach the final image plane appears capable, in principle, of obtaining spatial-intensity resolution of the type needed to detect certain extrasolar planetary systems. Perhaps the simplest and most direct method for canceling light from a star is to place an obstruction in the focal plane of the telescope, or at any plane in the optical train which is conjugate to the focal plane. This approach is used successfully, for example, in Lyot's coronagraph (ref. 27) for obscuring the solar disk. However, the Sun is an extended object – a star image is not. The very nature of a star image makes this procedure an engineering challenge. Most stars that would be studied in a search for extrasolar planets are so distant that their images will be unresolved. The object to be occulted therefore is the central region of the diffraction pattern, the so-called Airy disk. The size of the Airy disk depends on the aperture of the telescope and its focal length. Generally, the bigger the telescope, the higher its resolving power, the smaller the Airy disk and, therefore, the more difficult it is to make a device small enough and to locate it with sufficient precision to obscure it. It is conceivable that a telescope design could include an intermediate focal plane highly magnified so that obscuration could be accomplished there and this obscured image could be relayed to the system's final focal plane where it is detected.

Light that is blocked in this way is backscattered with much of it leaving the optical system. However, a small fraction of the backscattered light will "rattle" around through the optical system, ultimately arriving at the final image plane where it will either enhance the light in the Airy disk or, worse yet, appear as a spurious ghost image. In current applications utilizing central obscuration of this type, scattered light is not as severe a problem as it is for the low $\mathcal{R}$-value regime encountered in the planetary detection problem.

An alternative to the occulting disk approach is suggested here. This alternative makes use of the fact that light from a planetary companion to a star would not interfere with light coming directly from the star. The method is to line up the star with the axis of an interferometer so that the light in the two interferometer beams is recombined at the exit beamsplitter so that total destructive interference of starlight occurs. That is to say, the angle between the emerging wave fronts is so small that the distance between fringes exceeds the diameter of the exit aperture. Then an adjustment is made in the optical path length of one of the beams so that a dark fringe is centered on the exit aperture. Light from potential planetary companions would, if bright enough, be visible against this dark background. The basic concepts of such a dark-field device are discussed by KenKnight (ref. 28).

In order not to cancel out the planetary light while destructively interfering the two beams of stellar light, the interferometer is so constructed as to rotate one of the beams 180° relative to the other. One approach is as follows. The basic design (see fig. 24) is that of a Mach-Zehnder interferometer (ref. 29, pp. 312–315). Light enters the device at A (fig. 24), where it falls in a beamsplitter that divides the light into two approximately equal beams. One beam is reflected to mirror B where it is again reflected to the mixer located at D. The second beam passes through the beamsplitter to mirror C where it is also reflected into the mixer at D. Here the two beams are recombined. The mixer is structurally identical to a beamsplitter.

If all four elements are exactly parallel, the wave fronts incident in the mixer will be exactly parallel. However, if one of the mirrors, say B, is tilted slightly with respect to D, then the two wave fronts will not merge at the mixer but will be slightly inclined to one another. If the optical path lengths of two beams are equal, interference will occur and fringes will appear on the mixer. The direction of

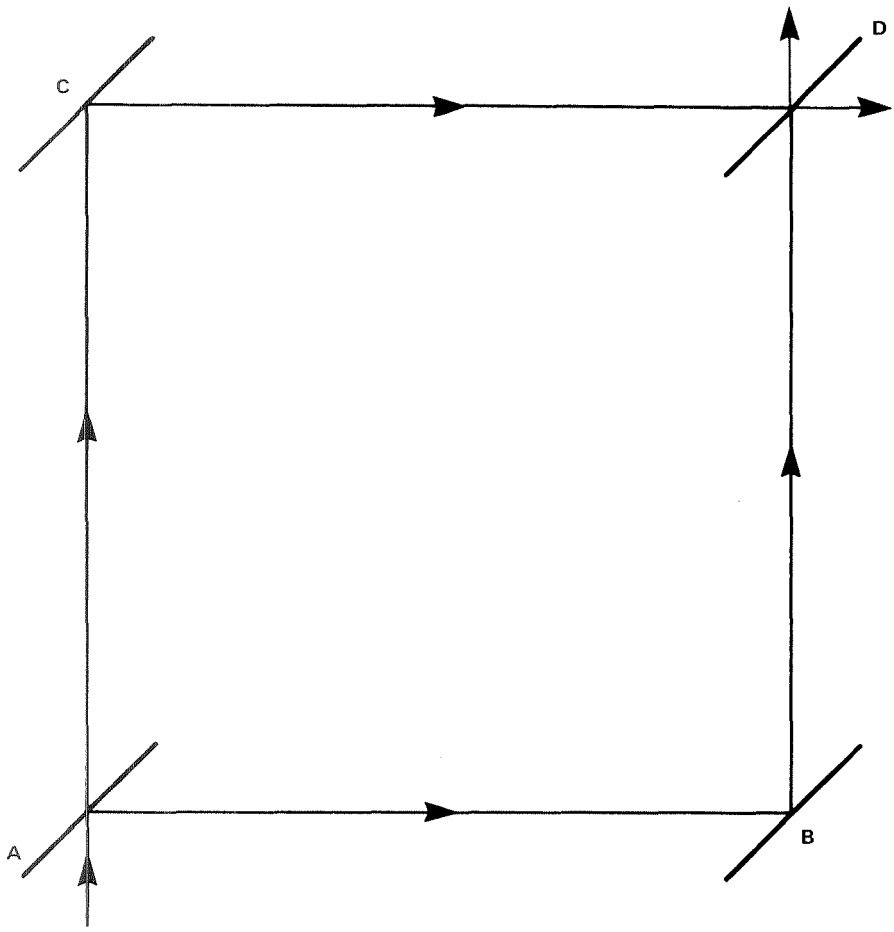


Figure 24.– *Schematic representation of a Mach-Zehnder interferometer.*

the fringes indicates the direction of the tilt; the separation of the fringes is inversely related to the magnitude of the tilt. This useful characteristic of the Mach-Zehnder is discussed later.

If the optical path length of one of the beams is increased, the fringes will move across the surface of the mixer, becoming fainter and fainter as the optical path difference approaches the coherence length of the radiation. If there is only a slight tilt between the two mirrors, the distance between the fringes will be large, perhaps larger than the width of the mixer plate. In that case, as the optical path

length varies, the entire mixer plate will vary in brightness from a maximum to a minimum. It is this principle that is applied to cancel the starlight without affecting the planetary light. A perfectly aligned Mach-Zehnder interferometer, with proper adjustment of the relative optical path length between the two beams, will null any entering wave front.

To null a wave front only when it is lined up with the axis requires rotating one beam 180° with respect to the other beam. The desired effect is accomplished by flipping one beam top to bottom and reversing the other beam left to right. This is done by replacing two of the plane mirrors in figure 24 with suitably oriented prisms (fig. 25). A wave front entering so that its normal makes an angle β with the axis of the interferometer will result in two wave fronts at the mixing plane making an angle of 2β with each other. With a perfectly aligned system, only a wave front with $\beta = 0$ will null itself. All

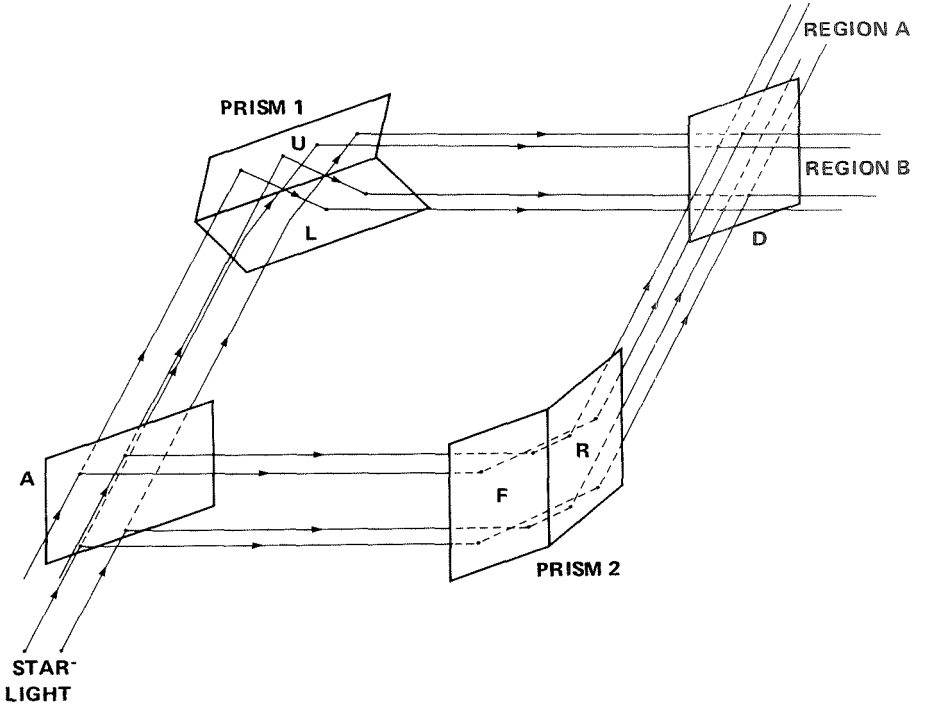


Figure 25.– *Schematic representation of the modified Mach-Zehnder interferometer discussed in the text.*

others will produce fringes whose separations are a function of 2β. As mentioned previously, the orientation of the fringes shows the direction of the tilt.

To evaluate the efficiency of the interferometer, it is useful to briefly review how interference occurs. The notation used is that of Ditchburn (ref. 30). The amplitude across the wave front $\xi(x,t)$ is given by

$$\xi(x,t) = a\, e^{i(\omega t+kx)} \tag{34}$$

where t represents time and x is a coordinate of the wave front. Interference occurs at the exit beamsplitter which makes an angle of 45° with the interferometer axis. To simplify calculations, it can be assumed that interference takes place on the exit aperture. The tangential components of the wave-front amplitudes on the exit aperture are

$$\left.\begin{aligned} \xi_+(x,t) &= a\, e^{i(\omega t+Ku \sin \beta)} \\ \xi_-(x,t) &= a\, e^{i(\omega t-Ku \sin \beta)} \end{aligned}\right\} \tag{35}$$

where u is a coordinate on the exit aperture in an appropriate direction. Summing these amplitudes gives

$$\xi = \xi_+ + \xi_- = 2a\, e^{i\omega t} \cos(Ku \sin \beta) \tag{36}$$

The fringe brightness will be at a maximum whenever $Ku \sin \beta = (2n + 1)\, \pi/2$ (integral n). The distance between adjacent bright fringes is given by

$$\delta_u = \frac{\pi}{K \sin \beta} \tag{37}$$

As β increases, the fringe spacing decreases to the extent that the fringes are unresolvable. On the other hand, as $\beta \to 0$, the fringes move farther apart until finally the distance between fringes exceeds the diameter of the exit aperture, and starlight cancellation becomes possible. If there is a fringe whose width is twice that of the aperture, then the entire aperture will appear bright with a 50-percent decrease in intensity at the edge. If the optical path length in one of the

beams is adjusted so that a dark fringe appears at the center of the aperture, then the entire field will appear dark with a 50-percent brightness at the edge.

Let d be the diameter of the exit aperture of the interferometer. Then this dark-field effect occurs when the distance between fringes exceeds $2d$, that is,

$$\delta_u = \frac{\pi}{K \sin \beta} = \frac{\lambda}{2 \sin \beta} > 2d \tag{38}$$

It follows that all wave fronts incident upon the interferometer at β, so that $|\beta| < \bar{\beta}$, where

$$\tan \bar{\beta} = \lambda/4d \tag{39}$$

will be interfered. Equation (39) may be inverted to determine d so that all wave fronts entering the aperture at the interferometer at angles less than $\bar{\beta}$ are to be annulled, namely,

$$d = \lambda/4 \tan \bar{\beta} \tag{40}$$

A telescope design that might be well suited to incorporating the apodizing and dark-field devices is a folded Gregorian, with several modifications. In the usual Gregorian, the primary is a paraboloid and the secondary, located beyond the primary focus, is an ellipsoid. In this design (see fig. 26), a tilted secondary convex spherical mirror is placed at or near the focus of the paraboloid directing the rays laterally toward an ellipsoidal tertiary mirror. This, in turn, reflects the beam of light to a focus. The light is intercepted before this focus by a diverging lens that acts as a collimator. Up to this negative lens, the telescope is an afocal system. The angular magnification M relates ray angles in object space and ray angles following the collimator as follows:

$$\tan \beta = M \tan \alpha \tag{41}$$

Here α represents the angle away from a star and β represents the off-axis angle of the corresponding ray in the region following the collimating lens.

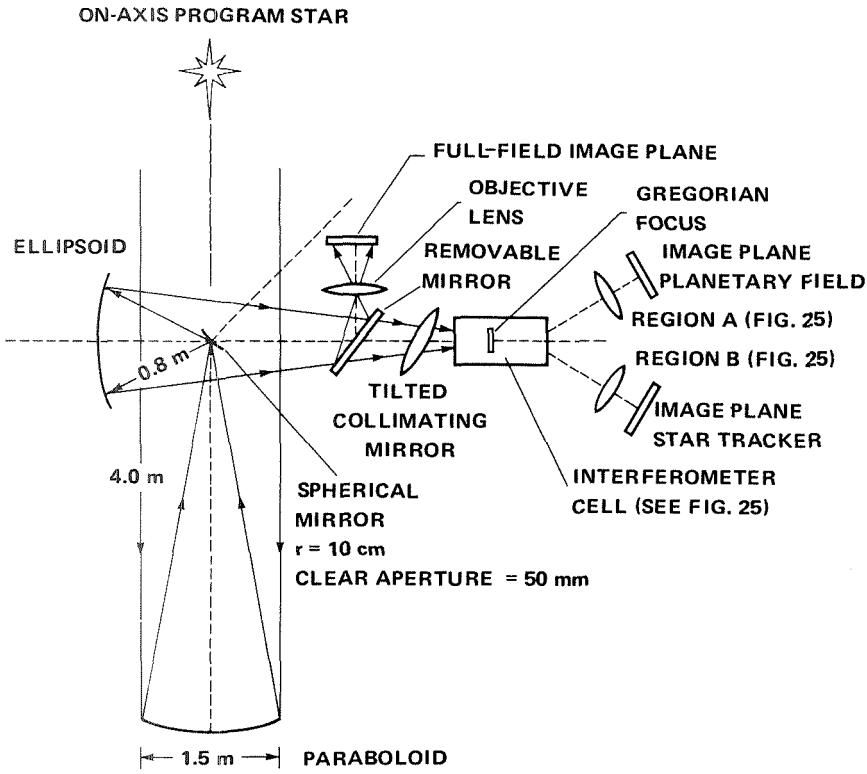


Figure 26.—*Schematic drawing of the modified Gregorian telescope considered for use with an apodizer and the modified Mach-Zehnder system (see text for detailed discussion).*

Equations (40) and (41) may be used to obtain an estimate of the exit aperture of the interferometer. The Airy disk, which is the portion of the diffraction pattern that the interferometer is designed to annul, has an angular extent given by $1.22\lambda/D$, where D is the aperture of the telescope. The post-collimator angular size θ^1 of the Airy disk is given by equation (41) as

$$\tan \theta^1 = M \tan \theta$$

However, θ and θ^1 are very small angles so that $\tan \theta \sim \theta$, or

$$\theta^1 \cong M\theta = \frac{1.22\, M\lambda}{D} \tag{42}$$

where, as discussed above, M is the angular magnification of the afocal part of the optics. To obtain cancellation over an angular subtent comparable to θ^1, the exit aperture must be

$$d = \frac{\lambda}{4 \tan \theta^1} = \frac{\lambda}{4\theta^1} = \frac{D}{4.88\,M} \tag{43}$$

Note that d is independent of λ (so long as the star is not resolved). For D = 1.5 m and a magnification M = 50, $d = 6.14 \times 10^{-3}$ m. An objective lens at or near the exit aperture of the interferometer would then image a dark field with doubled images of any off-axis objects (planets?).

A basic requirement for the system to achieve cancellation is that the interferometer optical axis be aligned rather precisely with the star under observation. This alignment must be achieved with a precision $\lesssim \theta^1$, namely, 4.07×10^{-7} M (in rad) (λ has been taken to be 5×10^{-7} m). Assuming that M = 50, $\theta^1 \cong 2.04 \times 10^{-5}$ rad $\cong$ 4 arcsec. That level of alignment precision is not overly demanding. Further, the system has a built-in alignment indicator. One attractive aspect of a Mach-Zehnder type of interferometer is that it can have two outputs. One may be used to obtain the image that will be analyzed for the presence of planetary companions. Output from a second beam is bright, with light from the star, and may be used to indicate the amplitude and direction of misalignment by noting, respectively, the separation of fringes and the direction of fringe tilting. This latter output may be automatically monitored and used to activate a servosystem that would align the optics to within the required tolerance.

The present study and analysis has not been carried to the point where it is possible to make a quantitative assessment of the extent to which the interferometer will provide a dark field. If it were capable of a reduction factor of $\gtrsim$100, the combination of the apodizer and interferometer would allow detection of systems with $\mathcal{R}$-values of $\sim 10^{-9}$ to 10^{-10}.

It is useful to make a few additional detailed remarks about the Gregorian design concept developed during Project Orion. The telescope system resembles a Gregorian reflector (see fig. 26). The primary mirror is a paraboloid with a 1.5-m diameter and with a focal length of 4.0 m. The diameter of the field at prime focus is therefore 35 mm. Located at the prime focus is a spherical mirror that operates a field lens. Its center of curvature lies on a line inclined 45° to the

axis of the paraboloid; it has a radius of curvature of 10 cm and a focal length of 0.05 m; and its diameter is 50 mm. Its axis is determined by the intersection of the axis of the paraboloid with its surface, and there are provisions to translate this mirror in two orthogonal directions for fine tracking.

The spherical mirror will produce a virtual image of the primary tilted approximately 24°. Since the dominant aperture of the system must be the primary, this image will be a pupil plane.

The third element is an ellipsoidal mirror. Its axis is 45° to the nominal axis of the spherical mirror and 90° to that of the parabolic primary. Its first (short) focus lies on the prime focus, a distance of 0.8 m from the ellipsoid. Its second (long) focus is at a distance of 2.4 m. Thus a beam of light is directed across the diameter of the primary. The ellipsoid's focal length is 0.6 m. The diameter of the field at the second focus is 3.1 m!

The ellipsoid will produce a pupil plane whose center lies at a distance of 1.936 m from the ellipsoid. This is an image of the primary mirror through the sphere and the ellipsoid. The pupil is tilted approximately 25.5° to the axis of the ellipsoid, and has a diameter of about 5.9 cm.

The next element in the optical train is a removable plane mirror tilted 45° to the axis of the ellipsoid. When it is in place, it diverts the full field to an objective lens where the light could be focused on a detector array. The full 0.25° field will cover the detector. Reference crosshairs will enable an operator to study the entire star field, to select a program star, and to point the telescope toward it. The removable flat mirror will be located 1.6 m from the ellipsoidal mirror; its dimensions will be 5 by 7 cm.

A lens at this point will form an image of the full 0.25° star field at its focus. Its field should be equivalent to the dimensions of the detector, that is, 3.6 m. These constraints are rather severe and, in the configuration described, the lens would resemble an inverted microscope.

Located at or near the pupil plane will be an achromatic collimating lens. If located 1.936 m from the ellipsoid, the required focal length of this lens will be −0.464 m. It will be tilted approximately 65.3° to the axis of the ellipsoid so that the pupil plane will be perpendicular to the axial ray of the telescope. The full aperture of this lens will be about 5.9 cm. The image of the pupil at this point is

reduced in size to about 9 mm. However, the semiangular field from the collimating lens is 65.3°. In practice, it would be better to shift the collimating lens slightly toward the ellipsoidal mirror, thus producing a real image of the pupil plane. It is at that location that the proposed apodizing element would be located.

The next item in the design configuration is the dark-field interferometer. Its function is twofold: (1) to cancel the starlight without affecting the light from the planet and (2) to detect how far and in what direction the star has drifted from the telescope axis and to provide a correction signal to adjust the position of the spherical mirror (which lies between the primary parabola and the ellipsoid).

Although the full-field output of the collimating lens is to be ±65.3°, only a portion of this need enter the interferometer. The spatial resolution required to detect a planet in the SPS (0.25 arcsec) translates into 0.07° in this space. The entrance aperture must be sufficiently large to accept a spread of only, say, 5° half-angle. The size of this pupil, as well as the detailed structure of the interferometer, were discussed above.

A lens at this point focused at infinity will produce two images of the planet in its focal plane. If the angular field of the rays emerging from the interferometer is, say, 5° half-angle, in which we wish to resolve an angle of 0.07° half-angle, a lens with a focal length of 41.1 mm is required. Its diameter will be kept small if it is moved up near the interferometer exit aperture. The field will now cover a radius of 3.6 mm to cover the 4- by 6-mm area of the detection device. At the other exit aperture of the interferometer, one will see a bright field when the star is properly aligned. If the star drifts off axis, fringes will appear. As noted earlier, because of the reversal and inversion of the wave fronts in the interferometer, the fringes will line up in the direction of the pointing error. These fringes are localized on the exit beamsplitter itself. A lens can be used to image these fringes onto the 4- by 6-mm field of the detection device.

It is not the intent of this report to suggest that this telescope design concept is by any means the optimal one for visual detection of extrasolar planets, let alone even a feasible concept in practice. To be sure, some of the quantitative aspects of the telescope appear to be absurd on the face of it. However, these absurdities can be greatly mitigated, if not altogether removed, by minor alterations in the design parameters.

In retrospect, several alternative procedures should be entertained. In this design, the tilt in the pupil plane, induced by the introduction of the spherical mirror, is corrected by tilting the collimating lens. It may be more expedient to tilt the ellipsoid instead. Here the advantage would be that the axis of the light emerging from the ellipsoid would be tilted so that the central obscuration caused by the spherical mirror and its accoutrements could be avoided.

DIRECT DETECTION AT INFRARED WAVELENGTHS

The portion of the electromagnetic spectrum that is loosely defined as the "infrared" lies in the wavelength interval $0.75\ \mu m < \lambda < 1000\ \mu m$ (1 mm). As with observations at visual wavelengths, turbulence in Earth's atmosphere effectively makes it impossible to search for extrasolar planets at infrared wavelengths from the ground. The atmosphere also presents problems in that molecules (e.g., H_2O and CO_2) in the atmosphere are very efficient absorbers of radiation over many regions in the infrared. These absorption effects can be minimized by taking observations from high-flying aircraft, such as the NASA C-141 Kuiper Astronomical Observatory (fig. 27), and they can be eliminated entirely by observing from space.

As noted in chapter 2, there are many appealing aspects of the infrared as a wavelength regime in which to search for extrasolar planets. Successful detection provides data concerning a number of significant parameters of the detected planet, most notably, temperature and diameter. In addition, the brightness ratio between star and planet (cf. fig. 6) is lower in the infrared than it is in the visual portion of the spectrum. However, detection of extrasolar planets by means of infrared (IR) observations is not without problems. This section discusses those problems and possible solutions.

An Infrared Interferometer Concept

It might be supposed that an IR telescope could rather easily have sufficient spatial resolution to detect the SPS. However,

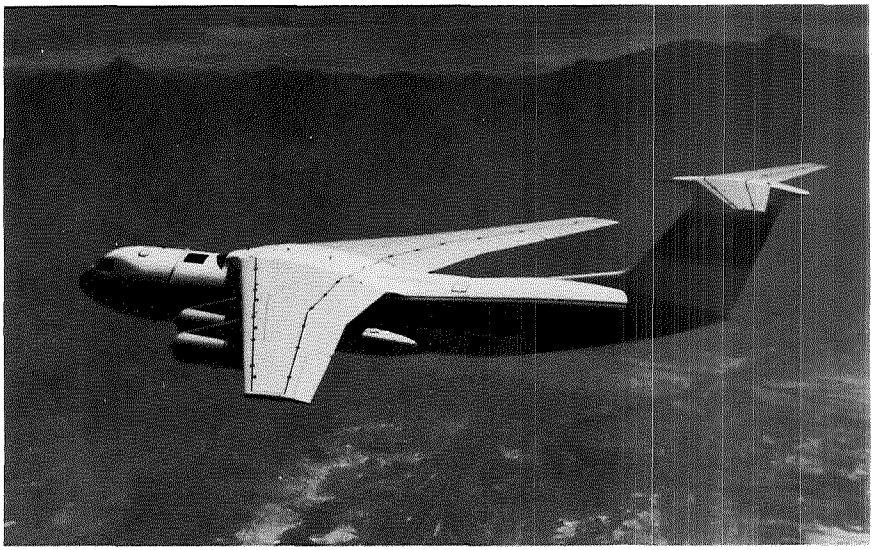


Figure 27.— *Photograph of the NASA C-141 Kuiper Airborne Observatory. The aircraft carries a 36-inch infrared telescope and permits astronomers to observe at altitudes above most of Earth's atmospheric water vapor.*

although the IR $\mathcal{R}$-value ($\sim 10^{-4}$) is indeed much more favorable than the visual $\mathcal{R}$-value ($\sim 2\times 10^{-9}$), it is still low enough to require a very large, clear-aperture telescope. Use of equation (32) for $\lambda = 40$ μm and $\mathcal{R} = 10^{-4}$ shows that the clear aperture required is $\sim$120 m. A further complication arises from the fact that, relatively speaking, apodization is less helpful at $\lambda = 40$ μm than at $\lambda = 0.5$ μm (because of the respective $\mathcal{R}$ values). Use of the Sonine functions as apodizing functions leads to a reduction in aperture size by about a factor of 40 at $\mathcal{R} \sim 2\times 10^{-9}$, but only by about a factor of 3 at $\mathcal{R} \sim 10^{-4}$. Thus, an apodized IR telescope must still be $\gtrsim$40 m. The approach taken in Project Orion to circumvent this problem was to consider an interferometer rather than a filled-aperture telescope. The angular resolution $\theta(S,\lambda)$ of an interferometer of baseline S is given by

$$\theta(S, \lambda) = \frac{\lambda}{2S} \tag{44}$$

Equation (44) describes the situation where radiation from a given source (a star) travels slightly farther, a distance of $\lambda/2$, to reach one

of the two apertures separated by distance S. The light amplitudes received at the detector must be added vectorially, giving rise to a null signal as the radiation at one aperture is 180° out of phase with radiation at the other aperture. Requiring that $\theta(S,\lambda) \sim 0.5$ arcsec at $\lambda = 40\ \mu$m gives $S \sim 8$ m, a reasonable size.

An interferometer has the added advantage of simultaneously providing angular and intensity resolution. If the star under study were a true point source, and if a space-based interferometer could be pointed with infinite precision *and* would remain absolutely stable, the stellar signal could be nulled out and the power from the two apertures balanced. Each of the "ifs" mentioned above is critical, and it is worthwhile to examine each in more detail.

Considering the SPS as a specific example, it may be seen that the star is not a point source. Its angular extent is $\theta_S \sim 0.001$ arcsec. The intensity pattern produced by the interferometer is of the form

$$I(\theta) \alpha \sin^2\left(\frac{\pi\theta}{\Theta}\right) \tag{45}$$

where θ here is the angle of the optical axis of the interferometer and Θ is the angle between successive maxima (or minima) in the $I(\theta)$ pattern. For $S \sim 8$ m, $\Theta \cong 1$ arcsec. The choice of $\Theta = 1$ arcsec places the SPS planet at the maximum of the intensity pattern and the star at the minimum (see fig. 28). The effective brightness ratio between planet and star is

$$\mathcal{R}_{IR} = \frac{KT_p}{T_*}\left[\int_p \sin^2\left(\frac{\pi\theta}{\Theta}\right) d\Omega \Big/ \int_* I(\theta_S)\ \sin^2\left(\frac{\pi\theta'}{\Theta}\right) d\Omega'\right] \tag{46}$$

where the interferometer intensity pattern is integrated over the two bodies, K is a constant, and $I(\theta)$ is the normalized intensity of the star as a line source. Taking the star image to be a uniformly bright disk of radius b, the integral over the star becomes

$$\int_*\left(1 - \frac{4\theta'^2}{\theta_S{}^2}\right)\sin^2\left(\frac{\pi\theta'}{\Theta}\right)d\Omega' = \left(\frac{2\pi}{\Theta}\right)^2 \int_0^{\theta_S/2}\left(\frac{\theta_S}{2} - \theta'\right)^{1/2}\theta'^2 d\theta' = \frac{\pi^3\theta_S{}^4}{64\Theta^2} \tag{47}$$

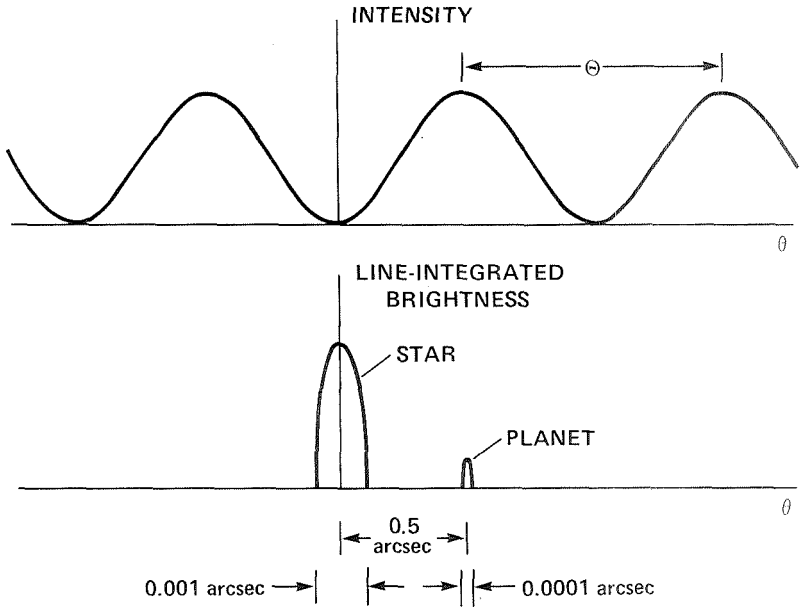


Figure 28.– *Schematic representation of the intensity of interference fringes for an interferometer, and the line-integrated brightness of a star and planet (not to scale) (see text for detailed discussion).*

The corresponding integral over the much smaller planet is simply the solid angle of the planet as seen from Earth. Thus,

$$\mathcal{R}_{IR} \sim \frac{KT_p}{T_*} \cdot \left(\frac{\Omega_p}{\pi^3 \theta_S{}^4} \right) 64\Theta^2$$

which gives $\mathcal{R}_{IR} \sim 60$ for the SPS; that is, the interferometer can, in principle, null out the stellar signal to the extent that the planetary signal is about 60 times *stronger* than the stellar signal.

Prior to discussion of the effects of pointing errors on the ability of an IR interferometer to detect extrasolar planets, it is necessary to make a few remarks concerning the operational aspects of such a device. In situations where small signals have to be detected in the presence of unwanted signals that might cause confusion, it is a well-established practice to modulate the desired signal. Consider the

previously described interferometer to be spinning with angular frequency ω about an axis passing through the star. Then the signal from the star would not vary in strength, but the signal from the planet would rise and fall with a fundamental frequency 2ω (fig. 29). The waveform would not be strictly sinusoidal, but of a characteristic flat-topped form $\sin^2[(\pi/2)\cos\ \omega t]$ containing a noticeable amount of the 4ω-harmonic (later shown to be 6 percent).

Very faint signals can be recovered, if they are modulated at a known frequency 2ω, by synchronous detection that filters out that frequency. Alternatively, where the modulation is substantially nonsinusoidal, the received signal can be broken into segments of duration $(1/2)\omega$ and averaged. The nonsinusoidal form may turn out to play a significant role. With this operational concept in mind, the effects of pointing error can be analyzed.

Effects of Finite Pointing Error

An element of the stellar disk at a distance e from the center of rotation and in position angle ϕ (fig. 30) will produce an output fluctuation proportional to $\sin^2[e\ \cos(\omega t + \phi)]$, and the total output due to the star will be obtained by integrating such contributions over all the elements of the disk.

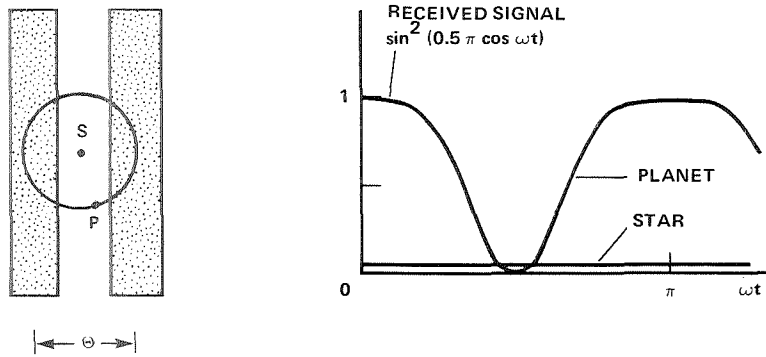


Figure 29.– *Schematic representation of the output signal from a planet (P) and star (S) as the planet traces a path through the fringe pattern of a rotating interferometer (see text). The planetary signal is flat-topped and rich in harmonics at a frequency four times the interferometer rotation frequency. A star centered precisely on the optical axis of the interferometer produces a small dc signal.*

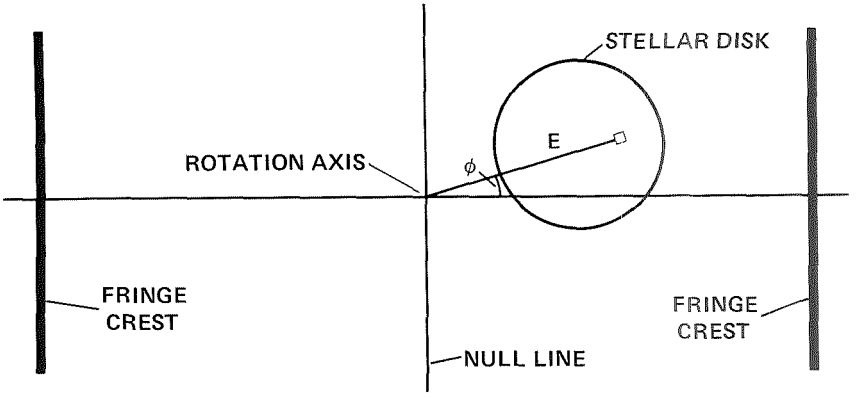


Figure 30.– *The geometry arising from a pointing error (E) in the optical system which leads to the null line of the interferometer not being centered on the star (see table 5 for numerical values).*

Instead of determining the waveform in its full detail, attention will be given to the peak value A_* and the trough value B_*. The configurations corresponding to peak and trough are shown in figure 31. Let E be the displacement of the center of the star from the rotation axis and let the radius of the star be R; then

$$\left.\begin{aligned} A_* &= \int_{E-R}^{E+R} 2\sin^2\left(\frac{\pi\theta}{\Theta}\right)[R^2 - (\theta - E)^2]^{1/2}\, d\theta \\ B_* &= \int_{-R}^{R} 2\sin^2\left(\frac{\pi\theta}{\Theta}\right)[R^2 - \theta^2]^{1/2} d\theta \end{aligned}\right\} \qquad (48)$$

The $\sin^2$ factors may be replaced by the squares of their arguments provided E and R are small compared with θ. The peak-to-trough amplitude is

$$A_* - B_* = \int_{-R}^{R} \left\{ \left[\frac{\pi(\theta + E)}{\Theta} \right]^2 - \left(\frac{\pi\theta}{\Theta} \right)^2 \right\} 2(R^2 - \theta^2)^{1/2} d\theta$$

$$= \frac{\pi^3 R^2 E^2}{\Theta^2} \tag{49}$$

The peak-to-trough amplitude of the variation due to the planet is calculated by giving the planet a radius $R/10$ and assuming that the received power falls to zero as the planet crosses the interference null ($B_p = 0$). If the planet had the same brightness as the star, the peak-to-trough amplitude would be $\pi(R/10)^2$, but, of course, the brightness is less by a factor of about 250. Hence we have as an approximate value for the amplitude ratio, when $E \ll \Theta$,

$$\frac{A_p - B_p}{A_* - B_*} \sim \frac{\pi(R/10)^2 \Theta^2}{250\pi^3 R^2 E^2} \tag{50}$$

Numerical examples of the ratio $(A_p - B_p)/(A_* - B_*)$ as a function of pointing error E are given in table 5. These results show that the spin axis of the interferometer may fall 4 stellar radii away from the center of the star before the unwanted amplitude of the 2ω variation due to the star builds up to equal or exceeds that due to the planet.

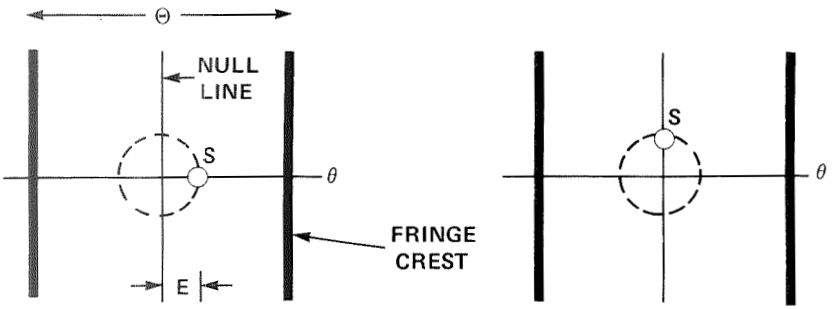


Figure 31.— *The geometry used to estimate the peak-to-trough variation in signal received from a star viewed with a pointing error E (see text for discussion).*

TABLE 5. POINTING ERROR EFFECTS ON THE RELATIVE MEASURED BRIGHTNESS OF PLANET/STAR

Pointing error E (arcsec)	Amplitude ratio
0.0005	16.7
.001	4.2
.002	1.0
.003	.43
.005	.16

It would not be surprising, although it would be technically demanding, if a passive object spinning in space could be guided to milliarcsecond accuracy on a star.

If the planetary signal could be distinguished from the stellar signal, then there would be no need for the amplitude ratio tabulated above to fall below unity. It would only be necessary for the received planetary signal to exceed some noise level, the nature of which will depend on instrumental design choices yet to be discussed. One possible design strategy (considered below) is to make the planetary modulation markedly different from the stellar modulation (in the presence of pointing error) by compressing the interference fringe pattern so that the planet is κ fringe spacings away from the star instead of half a spacing. This compression can be obtained by lengthening the interferometer baseline or by shortening the wavelength.

Under the arrangement previously discussed, the signal from the planet was of the form $\sin^2[(0.5\pi)\cos\ \omega t]$ but will now become $\sin^2[(\kappa\pi)\cos\ \omega t]$. The second harmonic content of this waveform is distinctly enhanced; in fact, it is possible to choose κ so as to suppress the fundamental component at 2ω completely. On the other hand, the signal due to the star, which is closer to the rotation axis, will be much more nearly sinusoidal with a fundamental frequency 2ω. Hence, by singling out the 4ω component, a frequency that is very precisely known, it should be possible to gain an advantage. One way of using this advantage would be to relax the pointing accuracy requirement on the spinning interferometer. Figure 32 shows that situation with $\kappa = 0.82$. As the fringe pattern rotates about the

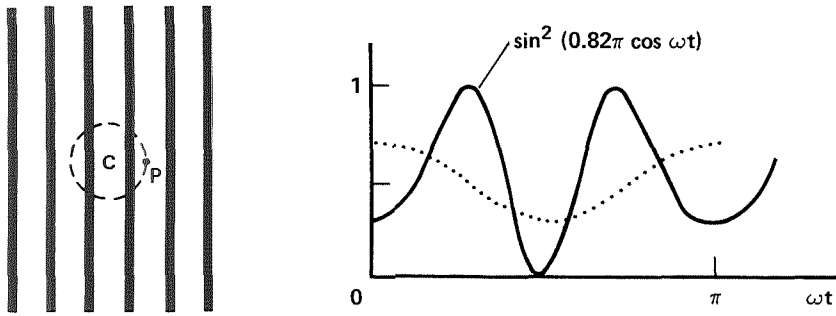


Figure 32.– *Schematic representation showing the effect of varying the fringe spacing relative to the angular extent of a planet's orbit. Spacing is adjusted to obtain a modulated signal optimized at a frequency equal to four times the rotation frequency interferometer.*

center C, the planet P (shown for $t = 0$) moves relative to the fringes around the dotted circle. The received power rises and falls as shown. This is the case mentioned above where there is no 2ω component at all. Meanwhile, if the star is off axis at a position corresponding to $\kappa = 0.5$, an angular pointing error of about 0.16 arcsec, it will deliver mainly a 2ω component and the 4ω component will be only 6 percent as large. Thus a considerable benefit derives from working with the 4ω component.

Before giving the basis on which the magnitude of the benefit may be calculated, it should be pointed out that another effect could contribute to distinguishing the planet from the star. There are even higher harmonics present in the waveforms, as is particularly obvious from inspection of figure 32, but the mix of harmonics is different for planet and star. Thus it will be possible to do better than would appear from comparing the 4ω components alone. All that can be done now is to demonstrate that an interesting possibility exists for relaxing pointing accuracy, but a careful quantitative study will require more effort.

The determination of harmonic content of the interferometer output waveforms is done as follows. To Fourier analyze

$$\cos(z \sin \omega t) = a_0 + a_2 \cos 2\omega t + a_4 \cos 4\omega t + \ldots$$

note that

$$\frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\cos(z\sin\omega t)\cos 2n\omega t\,\mathrm{d}(\omega t) = J_{2n}(z)$$

which leads to the results indicated in table 6.

These results indicate that for $\kappa = 0.5$, $a_4/a_2 = 6$ percent, as previously mentioned. The waveform $\cos(5.14 \sin \omega t)$, that is, $\kappa = 1.64$, roughly maximizes the amount of 4ω component and minimizes the amount of 2ω.

Noise Sources in the Infrared

There are three principal categories of potential noise sources for an IR planetary detection system: instrument-related noise, natural noise, and spacecraft-related noise.

Instrument-related noise can arise from both the detector and amplifier. The latter contributes Johnson noise, flicker noise, and shot noise. The former suffers from Johnson noise, current noise, temperature noise, and generation-recombination noise (ref. 31). The standard way to express noise level in IR astronomy is in terms of "noise equivalent power" (NEP). NEP (in units of $W/Hz^{1/2}$) is the minimum power that can be detected in 1 sec of integration time. Typically, good amplifiers have extremely low noise levels (NEP about 10^{-21} $W/Hz^{1/2}$). The noise level from a detector depends strongly on the type of detector, but IR detectors have been constructed with NEP's of 3×10^{-17} $W/Hz^{1/2}$.

TABLE 6. INTERFEROMETER HARMONIC CONTENT

	κ	a_0	a_2	a_4	a_6	a_4/a_2
$\cos(0.5\pi \sin \omega t)$	0.5	0.5	0.25	0.015	0.0003	0.06
$\cos(3 \sin \omega t)$	.95	.5	.49	.13	.011	.26
$\cos(4 \sin \omega t)$	1.27	.5	.36	.28	.05	.8
$\cos(5 \sin \omega t)$	1.59	.5	.05	.39	.13	8.0
$\cos(5.14 \sin \omega t)$	1.64	.5	0.	.39	.15	∞
$\cos(6 \sin \omega t)$	1.91	.5	-.24	.36	.25	-1.5

Among possible natural sources of noise are radiation noise from the source (planet), noise due to residual constituents of Earth's atmosphere, and noise due to zodiacal and background starlight. The NEP due to IR photon fluctuation from the planet is given by

$$\text{NEP} = [h\nu B(T_p)\epsilon A\Omega/B]^{1/2} \tag{51}$$

where h is Planck's constant, ν is the frequency (in Hz), ϵ is the emissivity of the emitting surface, A is the telescope aperture (in m), Ω is the solid angle subtended by the planet (in rad^2), B is the instrument bandwidth (in Hz), $B(T_p)$ is the brightness of the planet, and T_p is the temperature:

$$B(T_p) = \frac{2h\nu^3}{c^2} \cdot \frac{1}{e^{h\nu/kT_p} - 1} \quad \frac{\text{W}}{\text{m}^2 \cdot \text{Hz} \cdot \text{rad}^2}$$

For the SPS, observed with a telescope having $A = 1$ m^2, $B = 0.05$ Hz, $T_p = 128$ K, and $\nu = 7.5 \times 10^{12}$ Hz ($\lambda = 40$ μm), the NEP is about 10^{-26} W/Hz$^{1/2}$, a negligible effect.

Even at altitudes above 250 km, there are residual components of Earth's atmosphere, notably O, N_2, O_2, Ar, H, He, NO, N, CO, and the ionic forms of these species. The atoms and molecules that could affect an IR observation are O, NO, CO, and NO^+ (see fig. 33). The noise contribution from these species decreases rapidly with increasing altitude. An interesting feature of figure 33 is the absence of significant (10^{-17} W/Hz$^{1/2}$) IR radiation from the residual atmosphere between $\lambda = 6$ and 40 μm, the wavelength region of interest to a search for extrasolar planets.

Noise contributions from zodiacal light at $T = 304$ K and star light at $T = 5500$ K are also shown in figure 33. Zodiacal light is caused by interplanetary dust particles. From figure 33, it appears that zodiacal light produces noise dominating that due to the residual atmosphere. However, the zodiacal light curve shown in this figure obtains for the ecliptic plane, and the maximum intensity is at wavelengths below 20 μm. At a latitude $\pm 10°$ off the ecliptic plane, the intensity decreases by a factor of 3. At a latitude of $\pm 10°$ and at $\lambda = 40$ μm, the zodiacal light has a noise level of 3×10^{-17} W/Hz$^{1/2}$.

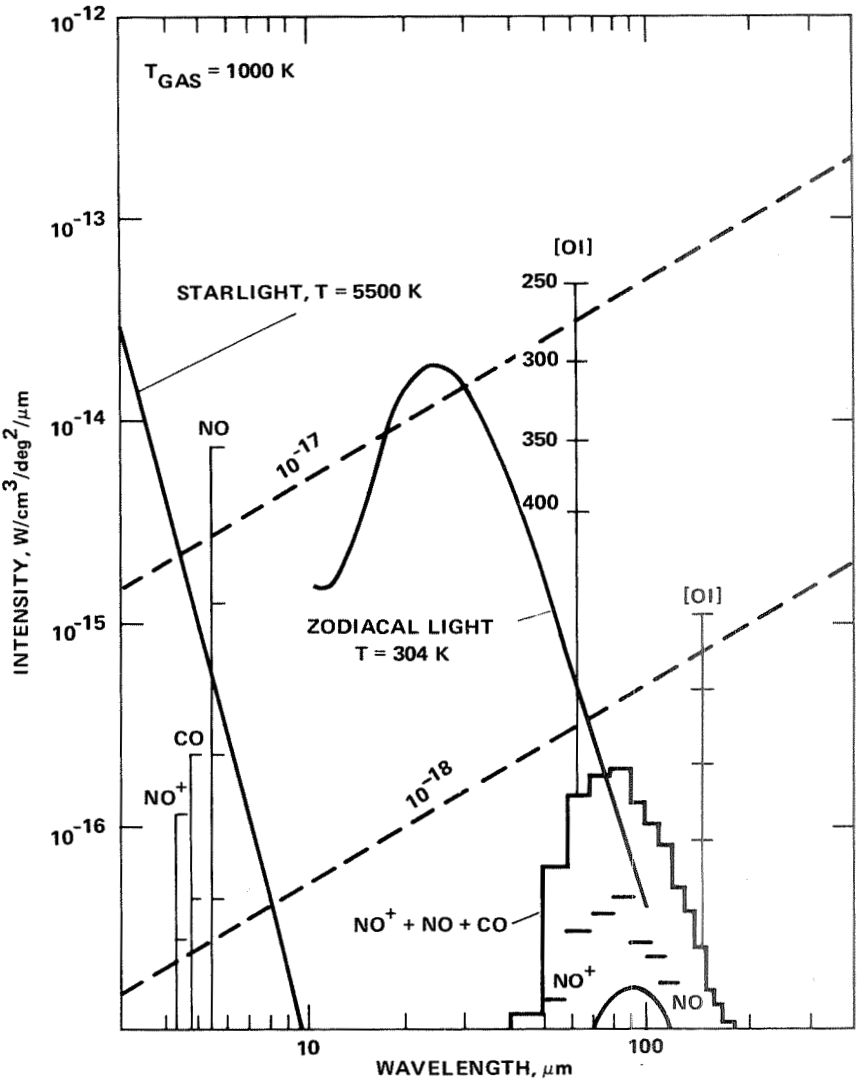


Figure 33.– *Potential sources of noise for infrared studies at an altitude of 250 km above Earth's surface. Sources of noise include residual neutral and ionized species of Earth's atmosphere, zodiacal light, and average starlight (figure from ref. 32).*

If the IR interferometer view is farther away from the ecliptic plane, then the zodiacal light should not be an important noise contribution if the wavelength is above 30 μm. However, in the range 5–30 μm, the zodiacal background could be a fundamental limitation of the IR space system.

The starlight shown in figure 33 is typical of a 5500 K source. The straight line is only the long wavelength portion of the complete curve which is similar to the blackbody radiation curve. Starlight is a troublesome noise contribution for $\lambda < 10$ μm. Only a small number of stars have significant influence on the background radiation above $\lambda = 10$ μm.

Moonlight, earthshine, and other planets in the solar system are other natural sources that increase the background radiation. Moonlight and earthshine are the two strongest and an IR system should be shielded from these sources to reduce background radiation.

One other potential source of noise for IR observations is that arising from effluent contaminants from a space shuttle or satellite. Two of the most important contaminants are H_2O and CO_2 (ref. 32). The high noise levels of both H_2O and CO_2 cover a wide range of wavelength spectrum. Even at an altitude of 400 km, the decrease of noise power as a function of altitude is rather slow compared to that of natural H_2O and CO_2.

Deposition of condensable gases (both natural and contaminants) on the surface of an IR telescope causes undesired absorption and scattering that degrades the performance of the IR space system. Because some of the contaminants generated by a space system have unusually high noise radiation, they should be minimized by: (1) choosing an orbit that requires a minimal number of convection maneuvers, (2) reducing the payload of the space system, (3) stabilizing the telescope and the interferometer rather than the shuttle or orbiter, (4) using low outgassing materials, and (5) storing waste water and venting it when not observing or when at poor observing locations. Items (2) and (3) are listed to minimize the required fuel.

Although a detailed analysis of the various noise sources discussed above has not been carried out, it appears that zodiacal light may be the principal contributor. It would indeed be ironic if the particulate debris within our own planetary system prevented us from discovering other planetary systems by means of IR observations.

Infrared Detectors

Development of technology in the area of IR detectors has been remarkably rapid. Any statement made here concerning the state of the art will certainly be passé at the time this report is printed. However, it is useful to delineate a few general aspects of IR detectors as applied to the planetary detection problem.

With the exception of the Golay cell, all of the effective IR detectors are solid-state devices. Any choice of an IR detector should involve the following factors:

1. Range of spectral wavelength
2. Required detectivity or sensitivity
3. Response time
4. Operating conditions (e.g., temperature and stability)

IR detectors can be divided into two main types:

1. Thermal detectors – the IR radiation is detected by measuring the change of properties or characteristics due to thermal effect. Some of the thermal detectors are thermocouples, thermopiles, Golay cells, bolometers, and pyroelectric detectors. Thermal detectors have a long response time, on the order of 1 msec.

2. Photoconductive detectors (or photon detectors) – these detectors utilize various internal photoeffects in semiconducting materials, such as a change in electrical conductivity or photoconductivity due to IR radiation. This type of detector has an extremely short response time, usually of the order of 1 μsec or less. This type of detector includes HgCdTe and PbSnTe detectors (intrinsic photon detectors), Ge and Si extrinsic detectors, MOSFET, and the Josephson detector.

The performance of an IR detector is based on the detector NEP (or the specific detectivity $D^* = \text{NEP}^{-1}$). An ideal IR detector (refs. 33 and 34) has the following NEP:

$$\text{NEP} = 4(\sigma k T^5 A \Delta\nu\eta^{-1})^{1/2} \tag{52}$$

where

σ Stefan's constant
k Boltzmann's constant
T absolute temperature, in K

A detector effective area
$\Delta\nu$ amplifier bandwidth
η emissivity of detector

In order to keep NEP low, the detector must operate at a low temperature, and the emissivity should be large ($\sim$1). Theoretically, the detector area should be small, but there are two limiting factors:

1. The practical feasibility of the detector – the detector cannot be very much smaller than the operating wavelength
2. The detector must be matched to the collecting optics, which requires that $A \geqslant \lambda^2/2\pi$ for the diffraction limited case

The bandwidth should be kept low, but not too low, to avoid losing the important signal. The broader the bandwidth, the lower the operating temperature must be for compensation.

SUMMARY

The problems attendant upon direct detection of extrasolar planets at either visual or infrared wavelengths are formidable. However, the potential gains are great, and the analyses carried out during Project Orion indicate that it may be possible to overcome these problems. Much more must be done, in far greater depth and detail, than was possible during Project Orion. More detailed studies on direct detection systems are being conducted by Stanford University, Hewlett-Packard, and Lockheed under the direction of D. C. Black of Ames Research Center. Hopefully, the preliminary efforts outlined here will provide a useful baseline for such future studies.

Thaw Measurement Refractor from Allegheny Observatory
Photograph © H. K. Barnett.

4. ORION IMAGING STELLAR INTERFEROMETER: A DESIGN CONCEPT

The principal effort of Project Orion was to develop a design concept for an improved ground-based astrometric telescope. The imaging stellar interferometer (ISI) design attempts to incorporate as many advantages as possible while avoiding most of the problems in various existing and suggested astrometric systems. As with a Miller long-baseline interferometer, the ISI ensures low systematic errors in relative angles because it samples light received from a given star at two widely separated points. As with a conventional telescope, the ISI compares the positions of numerous stars in a small area of the sky simultaneously, hence permitting correction of both atmospheric and instrumental parameters. As with several other telescopes built or proposed, the ISI incorporates a vacuum to improve instrumental stability and internal seeing. Unlike other interferometers, the ISI measures two relative azimuth positions. Relative zenith positions, which are subject to a large amount of continuously varying refraction, are not measured.

The ISI consists of two identical interferometers. Star positions are measured near azimuths that are 30° east or west of south (fig. 34). In operation, two tracking flats per interferometer rotate

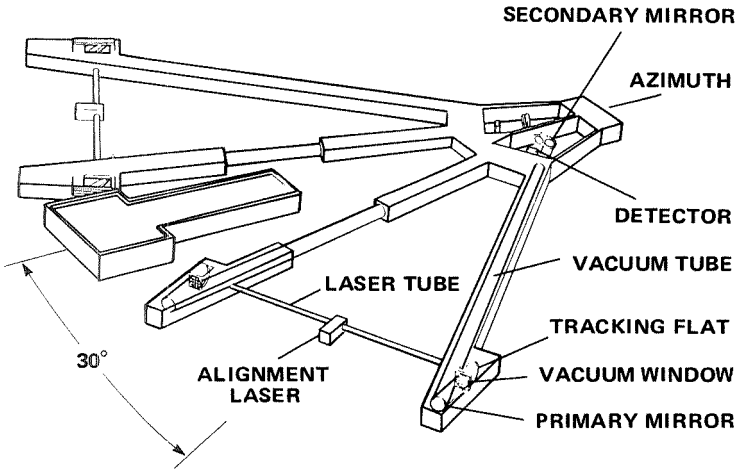


Figure 34.– *Perspective view of the Orion Imaging Stellar Interferometer (ISI).*

about a horizontal line to accept light from a desired zenith angle and each star is tracked vertically for approximately 5 min as it passes through the field of view. Stars first observed by the eastern interferometer are seen about 4 hr later by the western interferometer. Relative positions of stars are established by measuring relative X′ positions on one interferometer and relative X″ positions on the other interferometer, such as shown schematically in figure 35 for stars 1, 2, and 3 relative to star 0. Because the tracking flats rotate about a single axis and are the only moving optical elements, the system has a high degree of optical stability.

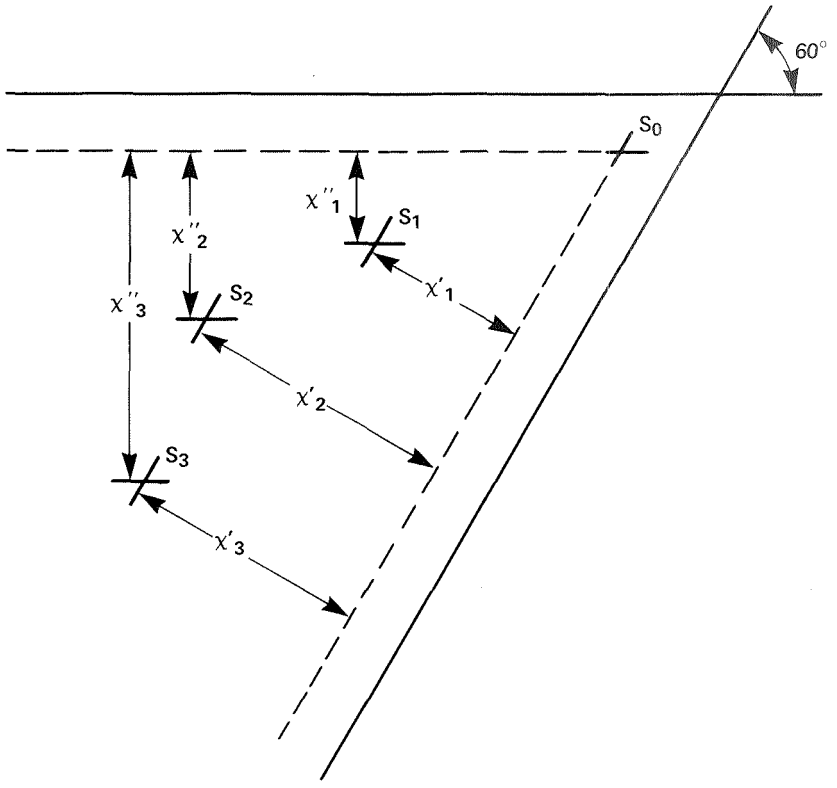


Figure 35.– *Schematic representation of the geometric construction used to determine the relative angular separation between stars using the Orion ISI (see text for discussion).*

In 1870, Michelson pointed out that if two small apertures are placed in front of a telescope, the light intensity in the focal plane is a "damped oscillation," such as shown in figure 36. The point of greatest oscillation about the mean intensity defines the apparent star position. In practice, the variable density of the turbulent atmosphere acts as a variable delay line, causing the apparent star position to move about its mean position. If a star's angular size is smaller than the resolution angle of the system, the intensity pattern will be as shown in figure 36. If a star's angular size is larger than the resolution angle of the system, the modulation envelope collapses toward the dotted line.

Although several interferometers operate on the above principles, the ISI is considerably different. The optics of the ISI lie in a horizontal plane, and light from a 1° by 1° region under study is reflected into the system by two tracking flats (fig. 34). The light

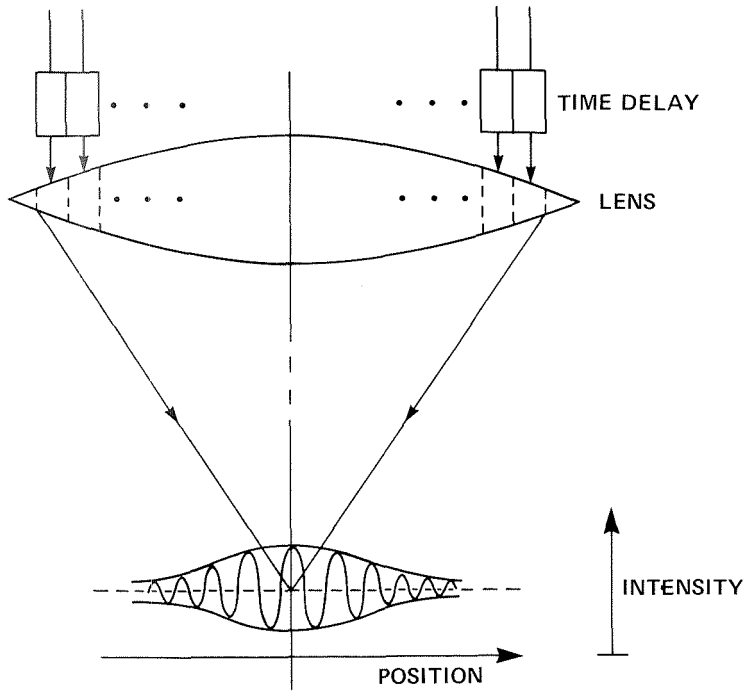


Figure 36.— *Schematic representation of a single-lens Michelson interferometer.*

then passes through segmented vacuum windows into the main optical system. The two primary mirrors are sections of what would be the primary of a f/2 Ritchey-Chrétien telescope with a 52.2-m aperture. Thus, they have off-axis paraboloidal surfaces that are slightly over 2 m in diameter. The secondary system is a figured flat consisting of two 1.85-m aspherical mirrors rigidly mounted next to each other (fig. 34).

The light reflected by the secondaries travels to an array of 20 detectors located a short distance behind the focal plane, as shown in figure 37. Each of these detectors tracks a designated star by moving along a system of ways (fig. 38). The position of the star is accurately and continuously monitored by measuring the position of the detector x with a laser system while the position of the star relative to the detector x is measured interferometrically as outlined below. The Y position of the detector is adjusted to track the star but is not used in the data reduction. The Z position adjustment is purely for focusing. An especially attractive aspect of this arrangement is that light from every star is gathered by the primary mirrors so that mirror errors are common to a star field. Yet detection is not in the focal plane so that all details of the mirror errors can be mapped. Imaging is used just to conveniently separate light from the stars in a field of view.

Within each detector (fig. 39) is a wave-front folding interferometer followed by a series of lenses, photomultipliers, and detector arrays. The wave-front folding interferometer (described in detail later) combines half of the light from the two input beams into each of the two output beams. The lenses and gratings further divide each beam into collimated beams of white, blue, yellow, orange, and red light. These beams form an image of one primary mirror and the mirror image of the other primary mirror on the photocathode of an image tube. The accelerated electrons from the images in each color impact a 16 by 16 array of charge coupled devices (CCD). Each element of the CCD array sees only a small portion of the primaries (approximately 0.4 percent) so that details of any errors on the primaries will be known. Mixing in the optical system results in each CCD monitoring the intensity signal from some point on the position scale in figure 36. A delay line scans that pattern; each detector sees a sequence of light and dark signals. If the greatest modulation occurs at a standard position of the delay line, then the detector is

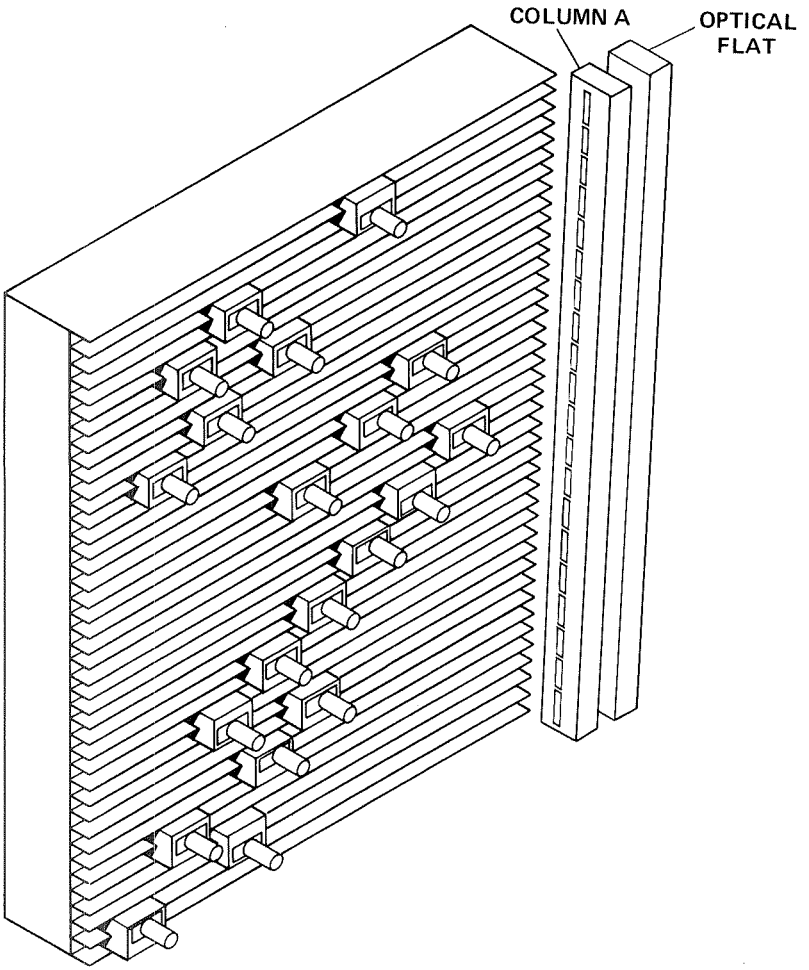


Figure 37.– *Schematic diagram of the detector plane assembly for the Orion ISI. Each individual detector (see fig. 38) tracks a given star by moving along the system of ways. The position of each detector is determined by laser ranging (see text).*

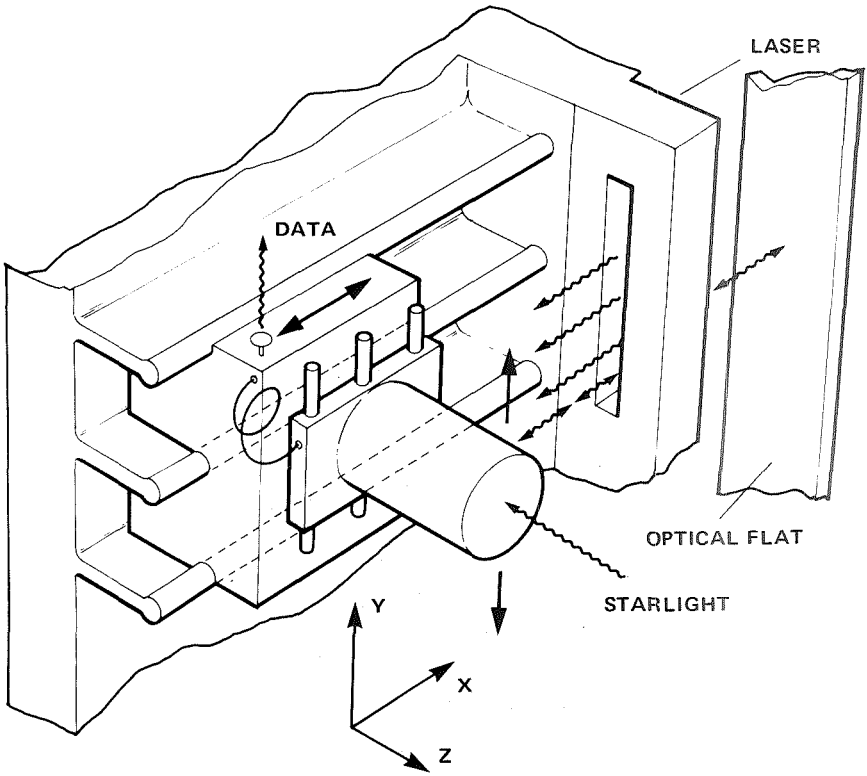


Figure 38.– *Schematic view of one of the 20 detectors mounted on the detector plane assembly (fig. 37). The principal detector motion is that of tracking (*x *direction). The required small motion due to rotation of the star field over the 5-min transit observation occurs along the ways in the* y *direction. A small adjustment in the* z *direction is needed to maintain focus as a function of* x *position.*

centered on the star and x is zero. The actual position of x for each CCD element is determined each 0.5 sec. The two output beams each give four separate color beams and each beam is detected by 256 CCD elements. We then obtain 2048 independent measures of x each 0.5 sec. These measures contain information on telescope and atmospheric aberrations. The former are mappable and highly constant and therefore removable. The latter are to be averaged out to produce relative star positions from data from each interferometer.

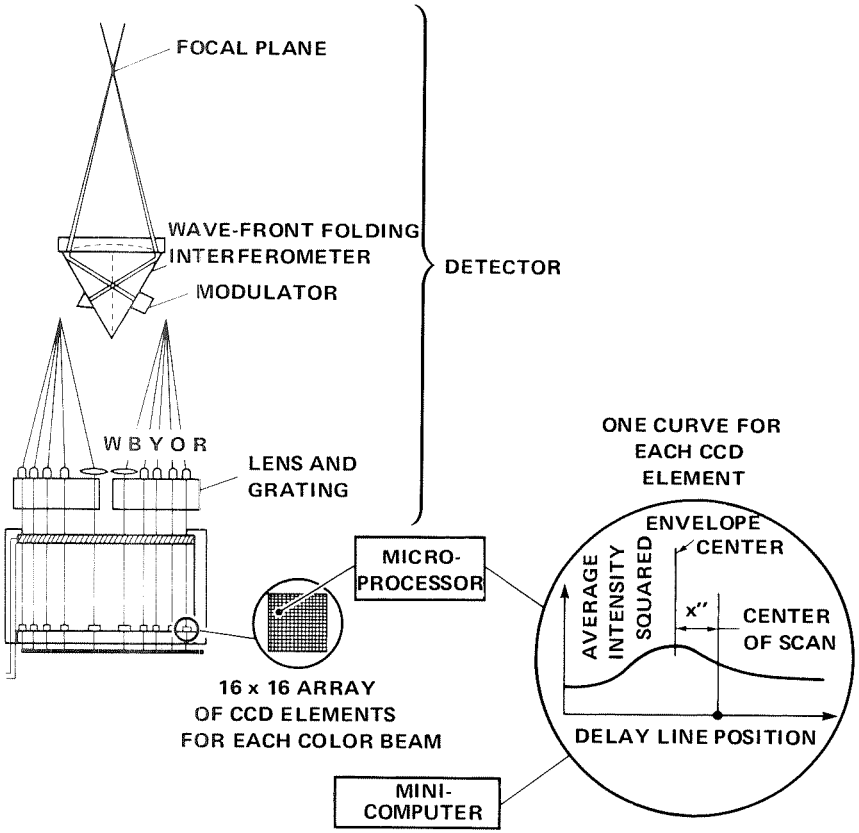


Figure 39.— *Schematic diagram of the optical system within a given detector. The figure shows the wave-front folding interferometer which produces white, blue, yellow, orange, and red fringe patterns. These patterns are imaged by a lens/grating combination on a photocathode which contains a 16 X 16 CCD array for each color. Representative output from each CCD element is shown at the right (see text for details).*

LIGHT-COLLECTING SYSTEM: ALIGNMENT AND GUIDANCE

The light-collecting system for the ISI consists of the optical elements, their mountings, drive, controls, vacuum system, and the superstructure that houses the instrument. Accuracy factors used to develop this system concept were established to maintain its inherent precision. This section discusses these accuracy requirements in terms of proposed designs.

Optical Requirements

The ISI consists of two identical imaging interferometers, one observing stars at azimuth $-30°$, the other at $+30°$. Each interferometer consists of two arms, eastern and western, each containing identical optical components, namely, a tracking flat mirror, a vacuum window, a primary mirror, and a secondary mirror. These components are described in the following subsections.

Tracking flat mirrors– Each mirror is elliptical in shape, with a major axis of 3.5 m and a minor axis of 2.2 m. The mirrors are rotatable around an axis that is parallel to the short axis of the ellipse and along the reflecting surface of the mirror. To maintain phase differences within detector requirements, the surfaces of the eastern and western tracking flat mirrors should be parallel within an angle of 0.1 arcsec. This accuracy is important because relaxing it by a factor of 2 decreases the visibility of the interference fringes almost to zero, making accurate measurements impossible.

The rotation axis should not deviate from the corresponding reflecting surface by more than 5 μm. To avoid realigning the tilt between the eastern and western tracking mirror whenever zenith distance is changed, it would be convenient to have these rotation axes parallel to the corresponding reflecting surfaces to within 0.5 arcsec.

Measurement of the relative tilt of the tracking flat mirrors is made possible by an evacuated pipe, 15 cm in diameter, joining the adjacent sides of these mirrors. A laser tracking system is installed in this pipe halfway between the mirrors. A small portion of the side

surface of each tracking mirror, facing the pipe and perpendicular to the main reflecting surface, is polished flat and aluminized to facilitate the laser tests.

The deviation from a perfect plane should not exceed 1 μm (for $\lambda \sim 0.5$ μm) over the entire 2.2 by 3.5 m surface of the tracking mirror. Any element of the surface, 15 cm in diameter, should not deviate by more than $\lambda/16$ from a plane parallel to the plane best fitting the entire mirror. Otherwise, fringe visibility for an individual element of the detector which "sees" an area of tracking flat about 15 cm in size would be noticeably degraded. The surfaces of the tracking mirrors, as well as those of the other mirrors, should be aluminized in such a way as to minimize polarization effects (ref. 35).

The thickness of each tracking mirror should be at least 30 cm to prevent random changes in shape in excess of $\lambda/40$. All mirrors in the instrument should be made of a material with a very low coefficient of thermal expansion, such as Cer-Vit or Corning ULE titanium silicate.

Vacuum windows— Vacuum windows serve as entrance pupils for the ISI. This is essential for performance of the instrument because stress birefringence in the windows would have different effects on each star in the field of view. The birefringent effects would vary with atmospheric pressure, temperature, and time. With the vacuum windows at the pupil, this effect is identical for all stars in the field and therefore has no influence on astrometric results.

Each vacuum window is 2 m in diameter. In order to be able to use a relatively inexpensive thin window, each window is divided into 0.25- by 0.25-m-square glass segments mounted in rigid steel frames. The detectors are aligned so that none of the individual detector elements "looks" at more than one square window element. Therefore, window elements need not have identical thicknesses. Nevertheless, no two of them should differ in thickness by more than 5 μm, nor should they be tilted relative to each other by more than 0.002 rad. Each element is paired with one in the other window by the detector scheme. Element pairs should be obtained from the same flat to make them alike in thickness.

The window elements should be made of glass having low stress birefringence, good resistance to climatic variations, and high transmittance. Grade-A Schott BK7 glass (or UBK 7 for better ultraviolet

transmittance), with birefringence not larger than 600 nm/m, seems to be most suitable. Each square should be polished flat to $\lambda/8$ and plane parallel to 0.2 arcsec.

Primary and secondary mirrors– Each individual detector subtends an area 12.5 by 12.5 cm on the entrance aperture of the interferometer. Increasing the linear size of this area is not practical because it would make tolerances on the relative tilt of the two tracking mirrors or two primary mirrors proportionately more strict. With this design, stars fainter than twelfth magnitude cannot be observed because the photon collection rate would be too low. Usable photon flux will be considered under the discussion on detectors.

As discussed earlier, about 20 reference stars brighter than fifteenth magnitude should be visible at any time. Thus, the region of the sky imaged by the ISI should be about 1° in diameter. This requirement, together with a practical diameter of about 2 m for the maximum primary and secondary mirrors, leads to the proposed dimensions of the ISI. The secondary mirrors must be 1.85 m in diameter in order to cover, without vignetting, an area of the sky 1° in diameter, assuming that the focal length of the primary mirrors (F_1) is 100 m. For this focal length, the longest baseline, that is, the separation of the primary mirrors, for which asphericities of primary and secondary mirrors are within the present state of the art, is about 50 m. Assuming that the diameter of the entrance aperture (vacuum window) is 2.0 m, the diameter of the primary mirrors must be 2.2 m.

The primary and secondary mirrors can be considered as portions of an imaginary, two-mirror aplanatic telescope with a single primary mirror of diameter $D = 52.2$ m and focal length $F_1 = 100$ m, that is, with f/2 focal ratio. To simplify the calculations of the shape of these surfaces, the fact that the vacuum windows are entrance apertures is neglected, and it is assumed that the primary mirrors act as stops. On this assumption, the formulas for eccentricities e_1 and e_2 of the primary and secondary mirrors of an aplanatic telescope given by Maksutov (ref. 36) and Gascoigne (ref. 37) may be used:

$$e_1 = \left(1 + \frac{2\beta^2}{\alpha - 1}\right)^{1/2} \tag{53}$$

$$e_2 = \left[\left(\frac{\beta + 1}{\beta - 1}\right)^2 - \frac{2\beta^2}{(\beta - 1)^3} \cdot \frac{\alpha}{\alpha - 1}\right]^{1/2} \tag{54}$$

where α is the ratio of F_1 to distance s, the distance from the center of the secondary mirror to the focus of the primary, and β is the ratio of s to the distance from the center of the secondary mirror to the final image formed by it on the axis. The radius of curvature at the center of the secondary mirror is

$$R_2 = \frac{2F_1}{(\beta - 1)\alpha} \tag{55}$$

The maximum deviation of each component of this imaginary two-mirror aplanatic telescope from a best-fitting spherical surface is

$$\delta = \frac{D^4 e^2}{512\, R^3} \tag{56}$$

Secondary mirror options– There are two principal options for secondary mirrors: (1) a secondary that is flat at the center ($R_2 = \infty$) and thus minimizes alignment problems, but is very difficult to manufacture, and (2) a concave secondary that is designed to be easy to manufacture. The first option will be assumed in later sections of this report.

1. Quasi-flat secondary mirrors: Let $R_2 = \infty$, which happens when $\beta = 1$. Moreover, assume that the secondaries are as close to the final image as is practical without obstructing the light beams by detectors. This minimum distance is taken as $s = 3.85$ m, which means that $\alpha = F_1/s = 26.0$. Equation (53) now gives the eccentricity of the primary mirror:

$$e_1 = 1.039 \tag{57}$$

The *imaginary* primary mirror of diameter $D = 52.2$ m is thus a hyperboloid of revolution, differing only slightly from a paraboloid (for which $e_1 = 1$).

For each of the *real* primary mirrors of diameter $D_1 = 2.2$ m, the radii of curvature are calculated at its edge, at the shortest and the greatest distance x from the axis of the interferometer. In the xz plane defined by the baseline and by the optical axis of the interferometer, the radius is

$$R_x = (4F_1^2 + x^2 e_1^2)^{3/2}/(4F_1^2) = \begin{cases} 204.64 \text{ m for } x = 23.9 \text{ m} \\ 205.54 \text{ m for } x = 26.1 \text{ m} \end{cases} \quad (58)$$

In the yz plane perpendicular to the xz plane and containing a normal to the surface at a given point, the radius is

$$R_y = (4F_1^2 + x^2 e_1^2)^{1/2} = \begin{cases} 201.54 \text{ m for } x = 23.9 \text{ m} \\ 201.83 \text{ m for } x = 26.1 \text{ m} \end{cases} \quad (59)$$

The surface of each primary mirror fits between the spherical surfaces of radii $R_x = 205.54$ m and $R_y = 201.54$ m. The maximum distance, Δz, between these two spherical surfaces over the surface of the primary mirror is

$$\Delta Z = \left\{R_y - \left[R_y^2 - \left(\frac{1}{2}D_1\right)^2\right]^{1/2}\right\} - \left\{R_x - \left[R_x^2 - \left(\frac{1}{2}D_1\right)^2\right]^{1/2}\right\} = 58\ \mu\text{m} \quad (60)$$

Thus the maximum deviation of the primary mirror from the best-fitting spherical surface equals $\Delta z/2$ or 29 μm.

An individual detector sees an area about 12.5 by 12.5 cm on a primary mirror. To secure good fringe visibility in the detector, the deviation of the mirror surface over this area from a perfect hyperboloidal surface should not exceed $\lambda/8$. Hence, over the entire surface of the primary, such deviation should not exceed about 4λ. The relative tilt of the two primaries (and, similarly, of the secondaries) should not deviate from the exact value by more than 0.2 arcsec.

If the secondary mirror were correcting only the spherical aberration and coma of the primary mirror, the maximum deviation δ_2 of its surface from the sphere best fitting both secondary mirrors simultaneously would be, from equations (54)–(56), for $\beta \to 1$,

$$\delta_2 = \frac{\alpha}{\alpha - 1} \cdot \frac{D_2{}^4}{2048\, s^3} \tag{61}$$

In this case, D_2 is the joint diameter of the two secondary mirrors. If each has a diameter of 1.85 m and the edges are spaced by 0.14 m, then D_2 = 3.84 m, which yields δ_2 = 1.93 mm. Asphericity of this magnitude is very difficult to achieve with sufficient accuracy. Probably, it is within the state of the art if a computer-assisted optical surfacing machine is used; such machines are presently in operation at Itek Co. and at Perkin-Elmer Co.

An individual detector sees an area ~5 mm in diameter on the secondary mirror. Therefore, over that area, deviations from the correct shape should not exceed $\lambda/8$ to secure good fringe visibility. The maximum deviation over the entire mirror surface should not exceed about 10λ or the detectors would operate too far from the white fringe.

The shape of the surface of a quasi-flat secondary needed to correct for spherical aberration and coma, in xy coordinates, is described by

$$x = \frac{\alpha}{\alpha - 1} \cdot \frac{y^4}{32\, s^3} \tag{62}$$

Here y is the distance from a point halfway between the two secondary mirrors. The surface of the secondary must be further deformed to correct for astigmatism. Using formulas given by Gascoigne (ref. 37), it is found that, for the ISI with a quasi-flat secondary, the astigmatism M is 1.35 times larger than for a single-mirror telescope of diameter D = 52.2 m. The deformation x needed to correct for this astigmatism is

$$x = \frac{MD^2}{16\, F_1{}^3}\, y^2 \tag{63}$$

where y is a distance from the line through the center of each secondary, perpendicular to the line joining the centers of both secondaries. In this case, the maximum additional deformation x needed to correct for astigmatism is x_{max} = 0.16 mm.

2. Secondary mirrors that are easy to make: This alternative should be chosen if the previously described secondary mirrors are found to be too difficult to manufacture. We assume

now that $e_2 = 0$, which is the case when a spherical secondary is sufficient for correcting spherical aberration and coma. For this case, equation (54) gives $\alpha = 13, \beta = 1.945$. This β is a factor by which the image of the sky on the detectors is demagnified compared to case (1). An area of sky 1° in diameter now requires a detector area only 0.90 m in diameter, separated by 3.96 m from a point halfway between the centers of the surfaces of the secondary mirrors. These surfaces are, as before, 1.85 m in diameter.

The eccentricities of the mirror surfaces are now

$$e_1 = 1.278 \; ; \qquad e_2 = 0 \tag{64}$$

The radius of curvature of the concave secondary mirror measured in the vertical plane is, by equation (55), $R_2 = 16.28$ m. In the horizontal plane, the radius of curvature is somewhat different because in this direction a deformation described by equation (63) is applied to correct for astigmatism. In this case, $s = F_1/\alpha = 7.69$ m and

$$M = [(2F_1/\beta) - F_1 + s]\beta^2/(2s) = 2.6 \tag{65}$$

Hence, by equation (63), asphericity of each secondary mirror is $x_{max} = 0.35$ mm. The asphericity is still large, but the surface of the secondary mirror is toroidal in shape, much more regular than in case (1). Inquiries at Itek Co. and at Perkin-Elmer Co. indicate that manufacturing such secondary mirrors with the required precision is well within the state of the art. Disadvantages of option (2) relative to option (1) are:

1. Greater sensitivity of secondary mirrors in option (2) to lateral shifts, which should not exceed a few μm
2. Image size on the detector is half that of case (1)

These disadvantages are not serious; therefore, option (2) remains an attractive alternative if manufacturing difficulties are excessive for option (1).

Tracking Flat: Support and Positioning

The success of the telescope depends largely on maintaining the surface shape and the rotational axis of the tracking flats regardless

of mirror orientation or ambient temperature. The viewing angles of ±40° from the local zenith and the primary mirror diameter of 2.2 m result in an elliptical tracking flat of 3.5 by 2.2 m. Figure 40 shows the tracking flat support system.

Tracking flat cell— The tracking flat is supported in a cell designed to minimize problems of gravity deformation, thermal expansions, nonuniform external loading, and inertia forces. The tracking flat cell defines the axial and radial locations of the mirror. It contains defining pads and a means of accurately controlling the force applied to these pads. Ninety-eight percent of the tracking flat weight is supported in the axial direction by air bags. Thus the load applied to the three defining pads is limited to about 445 N. Transducers will monitor the loads applied to the pads and will be interfaced with the air-supply system. Two air compressors will be used to ensure safe operation in the event of a pump failure. Similar pneumatic support systems (ref. 38) have been found to be extremely effective in eliminating mirror deformations while minimizing hysteresis and undesirable dynamic effects (ref. 39).

The tracking flat is supported radially within its cell by a mercury column contained in an oval tube and set in a groove in the mirror cell. The amount of mercury in the tube is adjusted until there is only a slight amount of radial motion permissible. Accurate radial positioning is then completed with adjustable defining pads

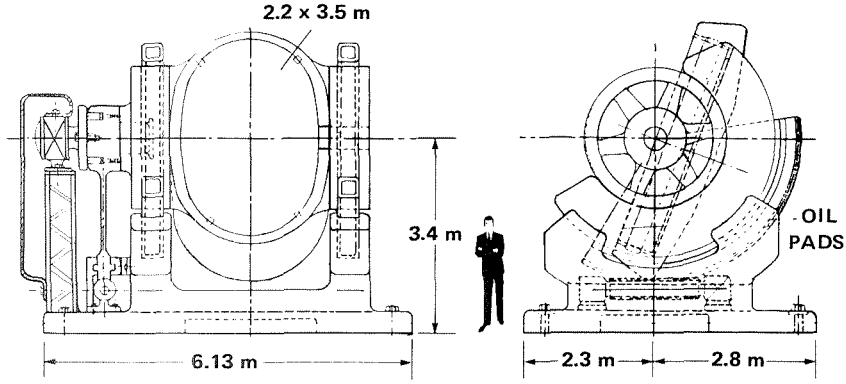


Figure 40.— *Schematic diagram of a tracking flat mirror assembly (there are four such assemblies for the Orion ISI).*

mounted in the mirror cell. Transducers mounted adjacent to the defining pads monitor the radial position of the tracking flat.

Support bearings– The tracking flat and its cell, weighing about 445,000 N, are supported on oil pads nestled in a cradle (fig. 40). The oil flows outward across the sill; it is filtered and then recirculated. A fluid film of 75 μm will be accurately monitored by transducers and maintained by the pump.

Preliminary calculations indicate that the maximum slewing friction at the bearings will be about 27 N•m, and the tracking friction will be about 0.23 N•m. Similar support systems have been used successfully with the Hale telescope and with the 1.55-m Naval Observatory telescope at Flagstaff (ref. 40).

Drive system– The drive system was designed to obtain maximum closed-loop precision in positioning, accuracy in tracking, and smoothness in operation. The accuracy of the drive system relies primarily on the gears and not on the servosystem. For this reason, the gear train selected was the one offering maximum accuracy. Other factors of less importance, such as efficiency and reversibility, were compromised.

Figure 41 shows the proposed power drive system gearing. Slewing is performed at a maximum rate of 4°/min, and the tracking rate is 0.051°/min. The tracking flat will accelerate to attain these speeds in about 4 sec.

A 4-N•m computer-controlled dc torque motor provides power for the tracking flat. A tachometer generator, attached to the motor, is used to obtain feedback. A torque-limiting clutch is included in the drive train to protect the gears in the event of a seismic shock or accidental collision. To minimize backlash, an auxiliary torque motor is used to constantly load the meshes to one side. The spur gear train contains three stages of reduction, with ratios of 5:1, 5:1, 5.6:1. The final and most important stage of the reduction is accomplished with a 4-m-diameter worm gear segment which has 960 teeth. It is driven by a single start thread worm having a 12-cm pitch diameter. In the past, design and construction of worm gears has been a troublesome area of telescope design (refs. 41 and 42); however, the present design solves most prior problems.

The worm is protected from impulsive-type loading by being mounted in a carriage that slides whenever the tangential force is too high. This carriage is mounted in a slide and held in position by a

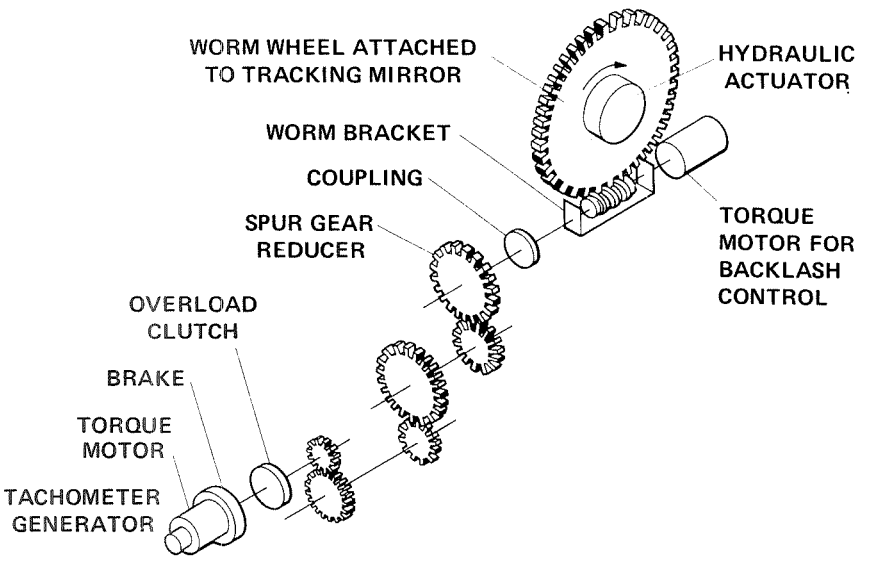


Figure 41.– *Schematic diagram of the power drive system for a tracking flat mirror.*

hydraulic actuator that controls the load transmitted by the worm teeth. The worm wheel is guided by hydrostatic bearings that guide the rim of the wheel so that the worm is maintained in constant contact. A surface-hardened worm mated with a phosphor bronze worm wheel should be used to minimize friction and wear.

Position control– Two control systems, both under computer control, are used to position the tracking flats. A hydraulic rotary actuator (fig. 42) will be used to counteract the gravitational unbalanced static torque acting on the worm gear drive. This counterbalancing reduces the power required by the dc torque motor and results in smoother, more accurate, positioning of the tracking flat. A dc torque motor drive (fig. 43) is used in two modes. First, the computer uses position control at the beginning of each tracking run to establish an initial position. The system then switches to velocity control and the angular tracking motion partially corrects for the rotation of the stellar field. The angular displacements of the two flats must remain synchronized to within 0.1 arcsec. The computer provides basic drive signals to the system and uses flat position data

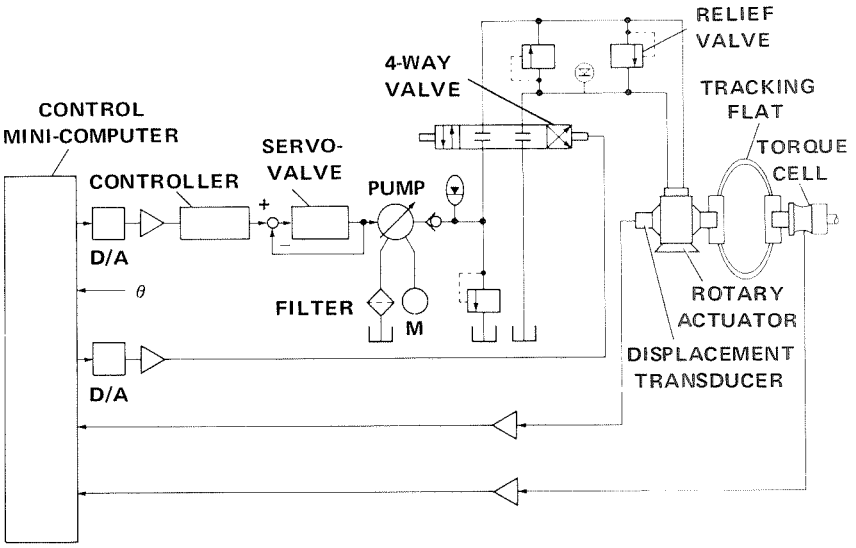


Figure 42.– *Schematic diagram of the tracking flat mirror torque actuator control system.*

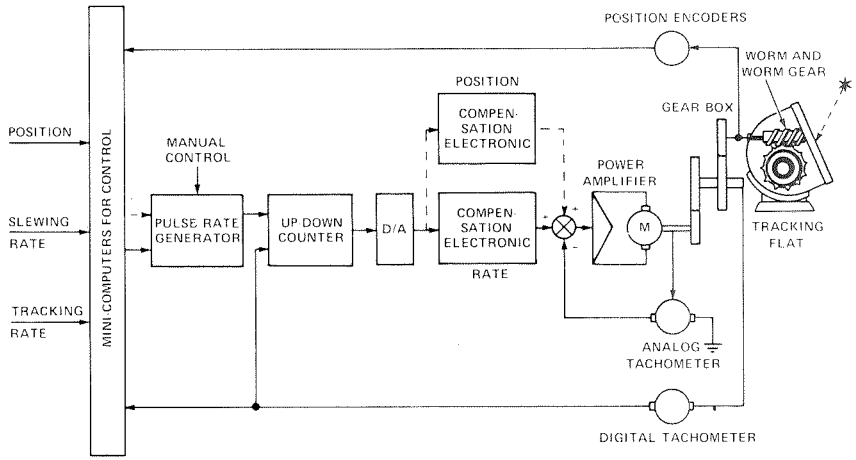


Figure 43.– *Schematic diagram indicating the tracking flat control system.*

to compensate for various system errors. The design specifications for the two control systems are given in tables 7 and 8.

Installation and alignment– Each mirror support will be mounted on a concrete foundation resting on bedrock (fig. 44). A leveled primary baseplate of steel will be mounted permanently on the concrete. Upon the primary baseplate will be a fine vertical adjustment plate (Z plate) overlain by two plates (X and Y plates) that will permit fine adjustment in two orthogonal horizontal directions. Each mirror support will attach to the Y plate.

Optical alignment of the tracking flat mirrors will be accomplished by a single axial alignment laser (fig. 45). A laser beam within an alignment vacuum tube will be split midway between the tracking flats into two nearly colinear beams aimed at the tracking flats. At the inner edge of each mirror, the lower half-circle of each beam will be reflected to a distancing detector midway between the mirrors. The upper half-beam will pass through a vacuum tube window with vertical reticle, then cross the reflective surface of the tracking flat

TABLE 7.– SPECIFICATION FOR TORQUE CONTROL SYSTEM

Gravitational unbalance torque (maximum)	1.8×10^5 in.-lb (2.04×10^4 N·m)
Hydraulic rotary actuator (double vane)	
Maximum torque at 3000 psi	71,000 in.-lb (8.03×10^3 N·m)
Displacement of hydraulic oil per radian	24.9 $in.^3$ (40.8 cm^3)
Unit weight	100 lb (445 N)
Angular travel	100° ±5°
Variable-volume hydraulic pump	
Delivery at 1200 rpm	9.3 gal/min (0.035 m^3/min)

TABLE 8.– SPECIFICATION FOR FLAT DRIVING CONTROL SYSTEM

Total angular displacement	40°
Tracking velocity	0.00085°/sec
Tracking acceleration	0.00045°/sec
Slewing velocity (maximum)	0.067°/sec
Rotational weight	13×10^4 lb (5.85×10^5 N)
Moment of inertia	1.13×10^5 kg-m^2
Motor (dc torque)	
Peak torque	3 ft-lb (4.1 N•m)
No-load speed	1500 rpm
Gear reduction	1.36×10^5
Encoder:	
Position	20-bit absolute fine 10-bit absolute coarse
Rate	20-bit/sec incremental
Control minicomputer	16K core memory 24-bit machine 10-msec sampling time
Tracking velocity accuracy	±0.001 arcsec/sec
Differential angular displacement during tracking	±0.10 arcsec
Positioning accuracy	±0.20 arcsec

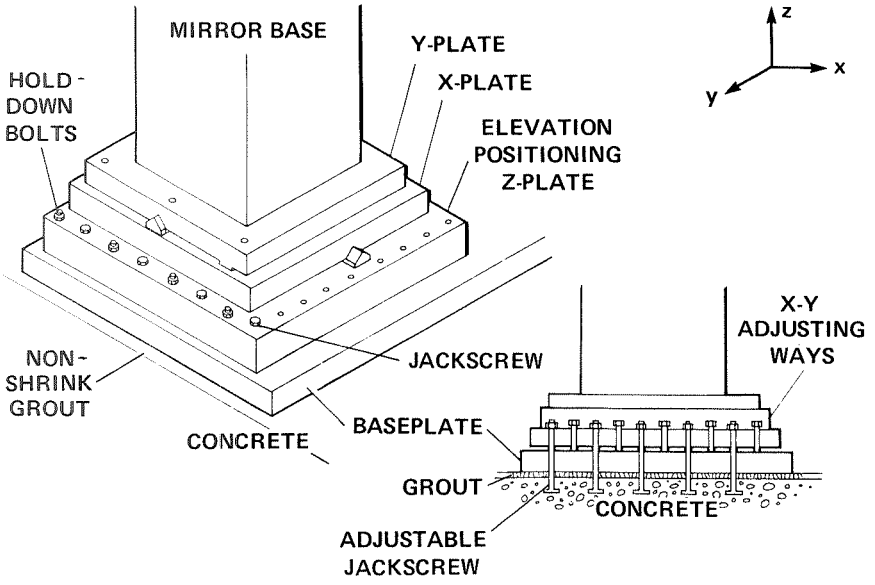


Figure 44.– *Schematic diagram of a mirror baseplate mounting.*

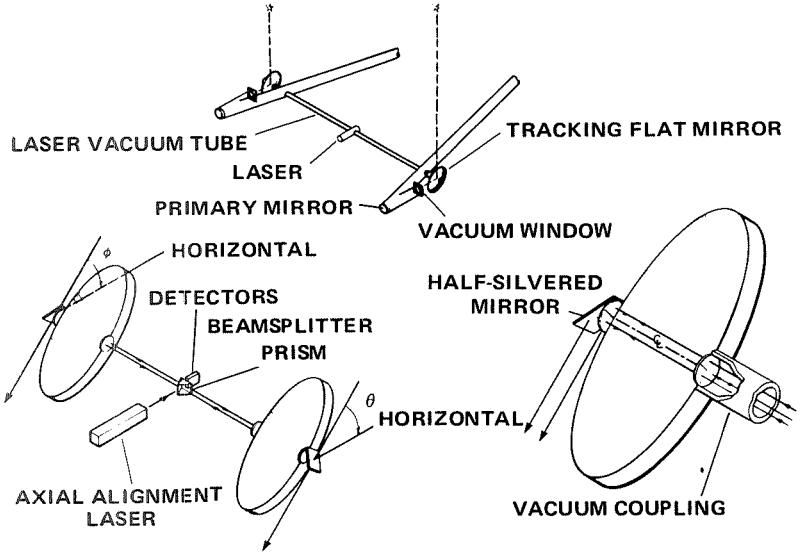


Figure 45.– *Schematic diagram indicating the manner by which a given pair of interferometer tracking flat mirrors are aligned (see text for discussion).*

mirror to a half-silvered mirror with reticle. Half of the remaining alignment beam will be reflected to a second detector or visual display midway between the mirrors. If the axial alignment of both large elliptical mirrors is colinear, the two beams and reticle images will coincide; slight misalignment will result in double images and significant misalignment will show no image. The final portion of the alignment beam passing through the half-silvered mirrors will be reflected orthogonal to the axis parallel to the reflective surface of the tracking flat mirrors. The angle this beam makes with the horizontal will indicate if the reflective surfaces of both mirrors are coplanar. If greater precision is desired, interferometric methods may be used with gratings on the vacuum window and half-silvered mirrors.

The actual angles (fig. 45) are determined by laser-distancing a horizontal side (x) and a vertical side (z) of a right triangle (fig. 46)

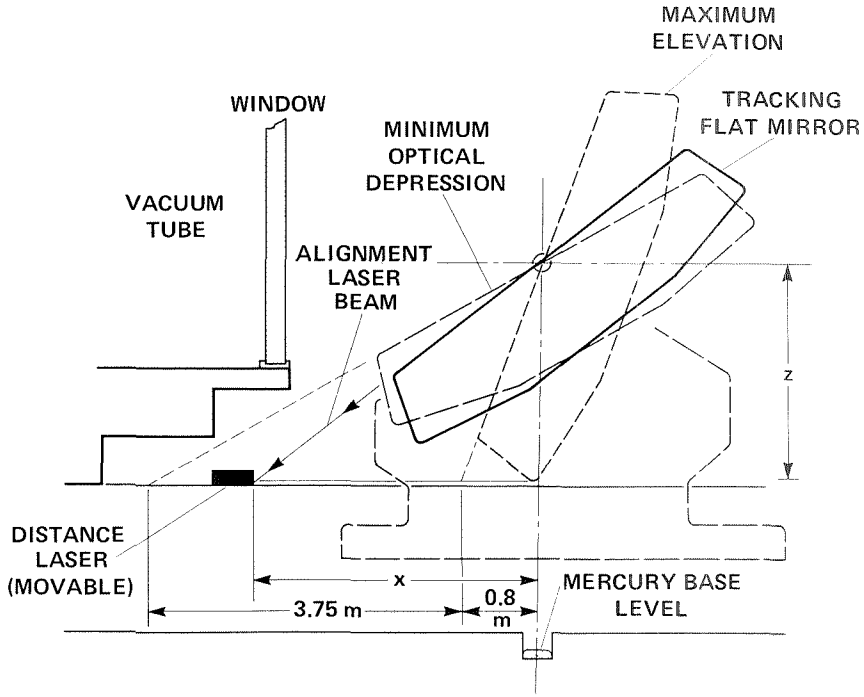


Figure 46.– *Schematic diagram of a tracking flat positioning system.*

(ϕ and θ = arctan z/x). If the angles are equal, no torque misalignment exists in the large elliptical mirror. A single distancing laser is proposed to determine x, which will vary, whereas z is a constant distance that needs to be measured accurately only periodically rather than constantly. Access to the mercury base level channel (fig. 47), which extends beneath the axis of each large elliptical mirror of the double interferometer system, will permit accurate placement and continued maintenance of the axes in a horizontal plane. The mercury base level channel should be semi-isolated from short-term atmospheric pressure differences. A small, optically reflective surface will be floated on the mercury at each observation point. Optical

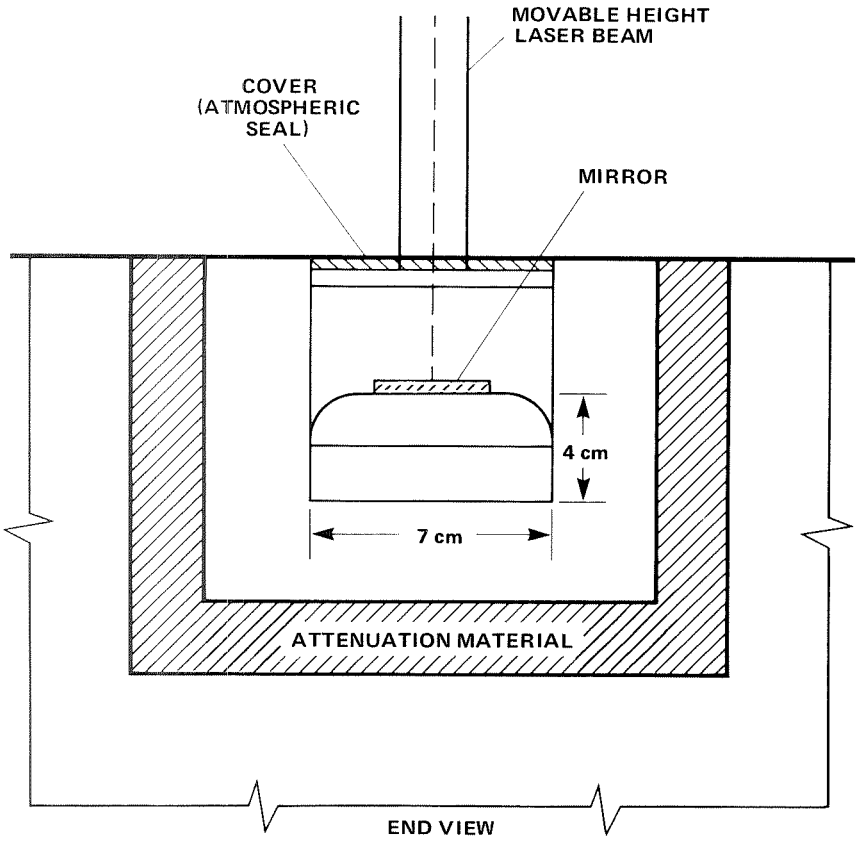


Figure 47.– *Schematic diagram of a mercury leveling channel (see text).*

alignment of the ISI is a more difficult problem because of the asymmetric optical elements involved. A laser system that permits high accuracy (ref. 43) could be used.

Enclosure for tracking flats– The tracking mirror is enclosed in a chamber. A folding door covering the top of the chamber slides open to expose the tracking mirror during observation. The opening allows the full light beam to reach the tracking mirror at all observational positions of the mirror. In the immediate surroundings of the opening, a windbreak, as well as a streamlined embankment, would be constructed to reduce wind disturbance (see fig. 48).

During the day and throughout the time period preceding observations, the folding door would be closed and the chamber would be maintained at the temperature anticipated during night observation.

Vacuum Window Support System

Figure 49 shows the assembly of the vacuum window, consisting of a square base frame with a circular opening and a circular mounting frame. In construction, the base frame is built in position

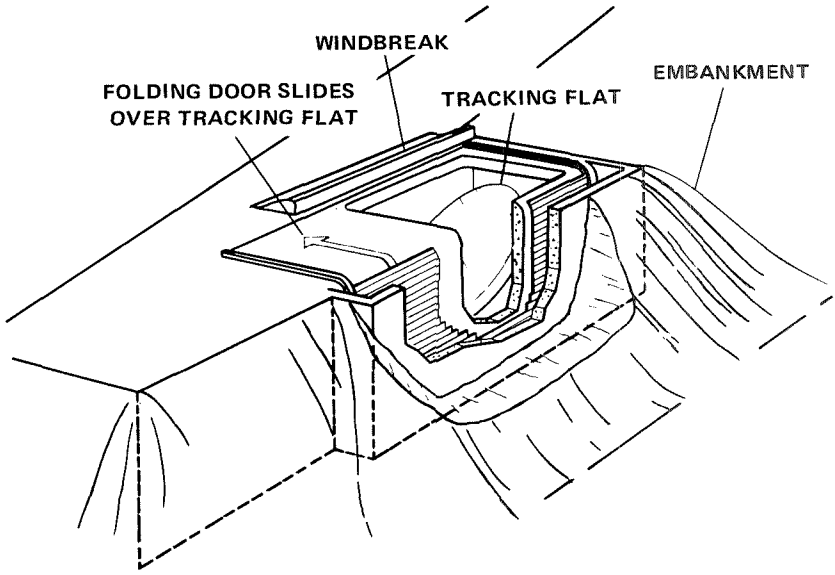


Figure 48.– *Cutaway view of a tracking flat enclosure.*

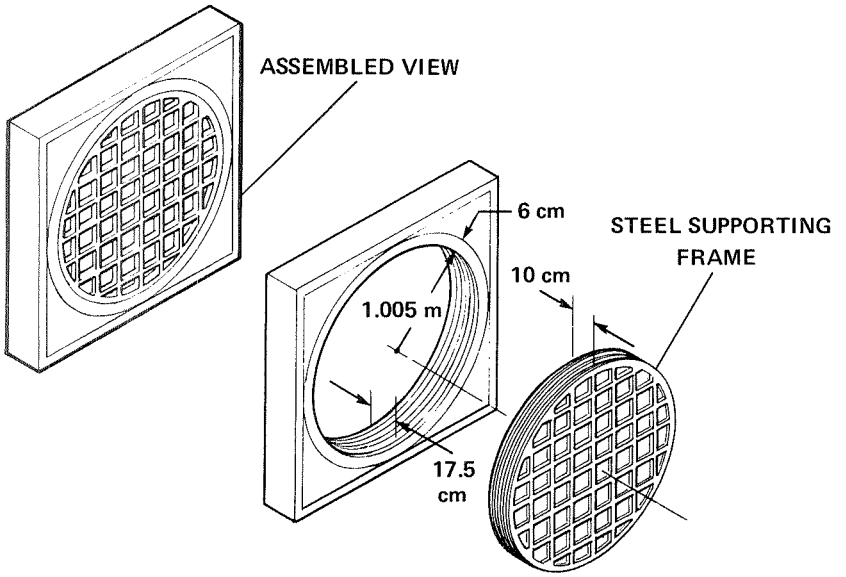


Figure 49.– *Schematic view of a vacuum window assembly.*

first, then the mounting frame is affixed to the opening of the base frame. The steel mounting frame has a grid system spaced at 25 cm in both directions. Fifty-two pieces of BK-7 glass of various shapes tailored to fit are mounted on the grid system. Rubber seals are used to prevent leaks. The maximum deformation and flexural stress of the 3-cm-thick glass sections are estimated at 7.6 μm (0.0003 in.) and 1.870×10^6 N/m^2 (270 psi). Approximately 9 percent of the optical area is obstructed by the grid system.

Primary Mirror Support System

The operation of the telescope requires that the primary mirrors located adjacent to the vacuum windows be segments of a paraboloidal surface with a 52.2-m diameter and a focal length of 100 m. Because the primary mirrors remain fixed in the gravitational field and are contained within the vacuum system, supporting structures can be made sufficiently rigid to ensure dimensional stability. A cellular structure can be applied to the mirror design to reduce

mirror weight and the weight of its supporting structure (as shown in fig. 50).

The primary mirror is positioned axially by adjustable supports that contact positioning pads on the rear surface of the mirror. A retainer on the front face of the mirror applies a small predetermined load to the locating pads. The primary mirror is located radially by a mercury-filled tube similar to those used to support the tracking flat. If necessary, edge-support locators can be provided to ensure proper alignment.

Secondary Mirror Support System

The secondary mirrors accept light transferred by the primary mirrors and focus it on a plane where it is measured and recorded by the detectors. Because the light falling on the primary mirrors is off-axis, aberrations of astigmatism and coma will occur. The secondary mirrors correct these aberrations.

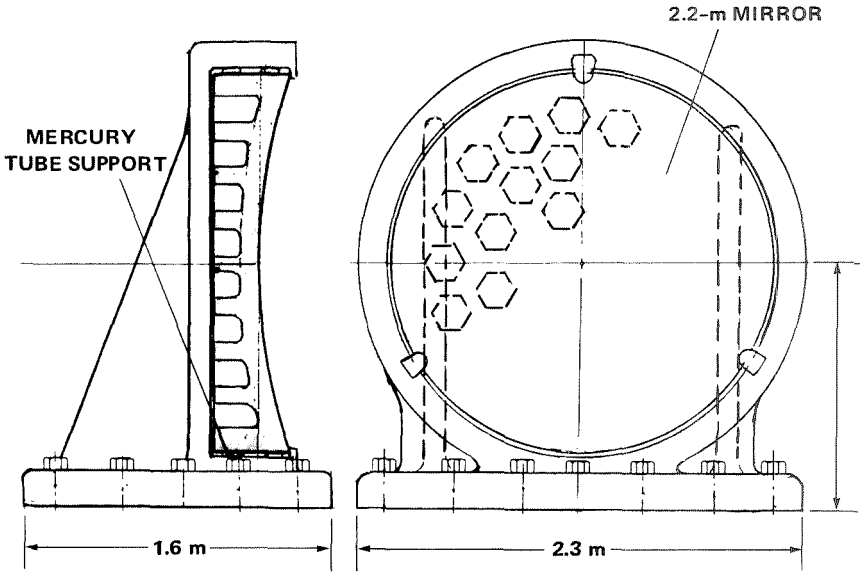


Figure 50.— *Schematic view of a primary mirror and its support system.*

To correct the aberrations, the optical surfaces of the two secondary mirrors should have a relative angular displacement of less than 6 arcsec. This implies that the rims of the mirrors should be controlled within a tolerance of 60 μm. Active control is accomplished by six piezoelectric actuators mounted on a stationary ring forward of each secondary mirror (fig. 51).

The secondary mirror cells are designed so that the mirrors float axially in air-support tubes around the mirror rim. When the actuators are retracted, the mirrors move forward against retainers on the cell. The two secondary mirrors are mounted on the same support. To minimize rheological deformations, the pair of mirrors and their inserts can be rotated 180° periodically.

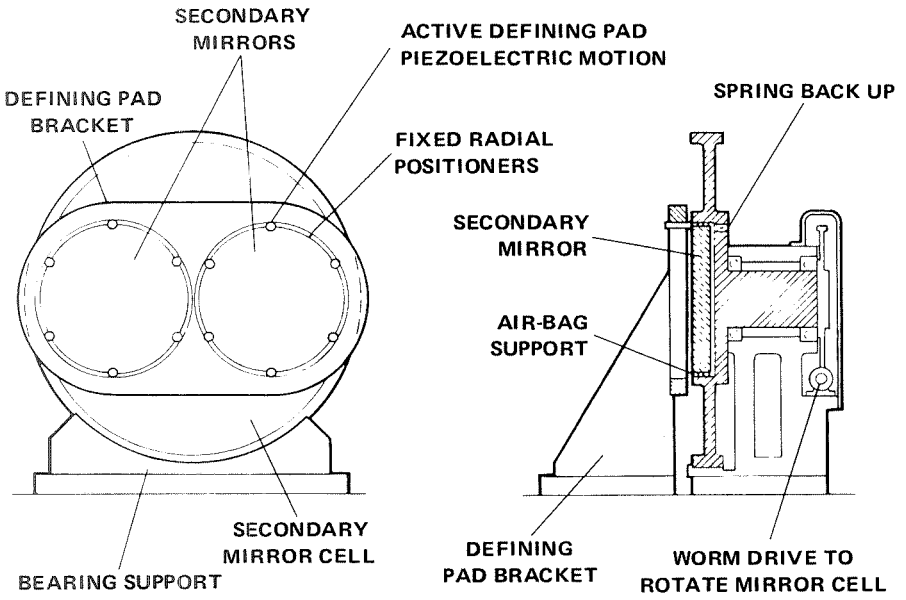


Figure 51.– *Schematic view of secondary mirrors and support assembly for one of the Orion interferometers.*

SIGNAL DETECTION, PROCESSING, AND TRACKING

Introductory Remarks

Detector philosophy– The essence of the suggestion by A. Michelson in 1890 was that some applications of a large telescope were simpler in interpretation if small regions around two points in the entrance pupil were the only regions allowed to contribute to an interference display. In contrast, the usual direct image is the consequence of interference between all possible point pairs in the entrance pupil. The latter display becomes complicated when the wave fronts coming from points in an object scene are distorted by the intervening atmosphere by more than 1 wavelength. But the distortions that are positive or negative with respect to the original wave front must have equal probability over a long time period. Therefore, an average wave front must approach the undistorted wave front in shape if the scene does not change appreciably during the averaging period.

Holography is an imaging procedure that registers the shape of a wave front for each point source in the object scene. Each point must furnish a reference wave front that interferes with the directly propagated wave front when they arrive at a detector surface. In Michelson's suggestion, the reference wave passes through the other of two small apertures in the entrance pupil. In the wave-front folding arrangement introduced in figure 39, the reference wave is again from another small aperture in the entrance pupil, but this time from a region obtained by reflecting the first region through a bisecting line through the center of the large synthesized lens. The reference region could be arbitrarily chosen elsewhere in the entrance pupil according to Michelson's suggestion.

The advantage of a wave-front folding geometry is that each possible point in the incident wave front is interfered with one and only one other point of that wave front. No information is lost by blocking off large regions of the incident wave front. The interference is displayed in an image of the entrance window. If the atmosphere were absent and the telescope optical elements were fabricated and aligned perfectly, that image would be uniformly illuminated by the light from one star, no matter how the star was

placed in the field of view of the telescope. That field of view is narrow, of the order of 10 arcsec, because the "eyepieces" on the moving platforms are small. Positional information in the field of view is revealed by the modulation of the ideally uniform display. If the star drifts across the field of view, the display intensity cycles from bright to dark and back to bright for an angular motion of λ/B for light of wavelength λ and separation of two points by distance B. The center of the field of view is defined by the cycle of greatest modulation, a definition independent of λ and λ/B. To avoid moving the star to the center of the field of view, that "center" is scanned across the star by a known modulation of the relative path lengths from eyepiece to detector surface. Thus, position data have been encoded into a temporal dimension.

In principle, we could form one large image of the entrance window following a wave-front folding arrangement of beamsplitter and mirrors. As in holography, we could combine the light from all stars in the detector plane. Two facts make such an approach impractical. The stars of interest vary in brightness by two orders of magnitude, so the dynamic range would be excessive. The stars more than a distance $Q\lambda/B$ from the optical axis give interference fringes of negligible depth, where $Q = \lambda/\Delta\lambda$ is defined by the wavelength bandpass $\Delta\lambda$. For $Q = 15$, this angle is only 10 marcsec at $\lambda = 0.5$ μm and $B = 50$ m. Thus the design requires independent detector systems and synchronous modulation within each moving eyepiece.

To understand the consequences of misalignment for position measures, consider various tilts. The tracking flats should form one plane. If the atmosphere were ignored and the only tilt error in the telescope were a relative rotation of the normals to the flats about the horizontal tracking axis x by θ_t, then the wave fronts at the exit pupil would be tilted by $2\theta_t$. The tilt causes an apparent trend in star position proportional to detector y position measured from the window centerline:

$$\Delta x = 2y\,\theta_t \tag{66}$$

Such a linear trend drops out in the average over all detectors in the aperture, but will cause serious loss of modulation depth at every detector if the linear trend exceeds $\lambda/4$ in a distance d across each detector. Thus, by setting $y = d$,

$$\theta_t < \lambda/(8d) \tag{67}$$

In the present design, $8d = 1$ m, so $\theta_t < 0.1$ arcsec. Similarly, all horizontal and vertical tilt errors of tracking flat, primary, and secondary mirrors should have a sum that does not exceed θ_t.

As might be expected from our holographic analogy, the system is insensitive to focus of the collimating lens at the entrance to each detector. Let the primary telescope have focal length F_1 and the collimating lens have focal length F_2. Suppose the focal planes are separated by distance e along the optical axis. After the collimator, a plane wave front would then be deformed by

$$\left.\begin{aligned} W_1 &= b(r - r_1)^2 \\ b &= e/(8F_1^2) \end{aligned}\right\} \tag{68}$$

if we consider the wave front projected onto the primary. The copied and inverted version of the wave front arising from the second arm of the interferometer has a deformation centered on a different point, generally,

$$W_2 = b(r - r_2)^2 \tag{69}$$

The intensity distribution at the image tube depends on the path-length difference which includes the term

$$\delta = W_1 - W_2 = b(r_1^2 - r_2^2 - 2r_1 r + 2r_2 r) \tag{70}$$

As before, we are concerned with how rapidly this changes in distance d, so we differentiate δ with respect to x and multiply by d:

$$d \cdot \delta' = 2bd|r_1 - r_2| \leqslant \lambda/4 \tag{71}$$

Note that $r_1 = r_2$ if the star is on the axis of the collimator. The distance that the star is off-axis in the region of appreciable output modulation is about Qfd, where $f = F_1/B$ is the focal ratio of the system. Projected on the primary, we find that

$$|r_1 - r_2| = Qf\lambda F_1/F_2 = QF_1^2\,\lambda/BF_2 \tag{72}$$

Finally, the requirement on focus (apart from light missing the detector lens) is

$$e \leqslant \frac{B}{d}\frac{F_2}{Q} \sim 3\ \mathrm{m} \tag{73}$$

To keep the light paths constant inside the detector, the focus range is 3.8 mm from the edge of the field of view (1.75 m diameter) to the center.

We could continue to describe how various other wave-front perturbations become visible in the interferometer display. The specification of the optical system tolerances which follows reflects the requirement that those perturbations remain small. The essential point is that a detailed data analysis scheme will permit the measurement of important system perturbations. If the perturbations are acceptably small, they may be removed during reductions. If not, they form the basis for improving alignment.

Interfacing with the ISI– We noted earlier that the light from a star in the center of the field of view travels in a horizontal plane after reflection from the tracking flats. The light proceeds to a primary mirror and a secondary mirror before it reaches the detector plane. There the rotation of Earth requires that the tracking flats turn in elevation at a constant rate to keep the central starlight in a horizontal plane. The star images at the focal plane move at a slightly variable rate in an essentially horizontal direction when the central star moves horizontally. Star detectors move horizontally at a programmed rate to follow the stars across the field of view.

Position data are to be obtained at several colors simultaneously for each star. In principle, one color is enough because the horizontal component of the angle between two stars is independent of color if the atmosphere is horizontal. The vertical component is complicated by dispersion in the air prism over the telescope. That dispersion amounts to an angle of the order of 1 arcsec from blue to red light at moderate elevation angles. But surfaces of constant density in the atmosphere may be inclined by a few milliradians due to weather fronts. Thus the horizontal component may be disturbed at the milliarcsecond level, and such disturbance will be revealed by comparison of the position data at several colors.

A schematic suggestion for mounting traveling microscopes on horizontal rails was shown in figure 37. The microscopes are, in

effect, sensitive optical tiltmeters, so the supports must be rigid and smooth. The number of detectors is set by the need for at least five reference stars that will prove suitable over a decade of research on each selected sky area. Current suspicions based on stellar radial velocities suggest that 50 to 75 percent of all stars will have companion stars with periods in the 10- to 100-year range. Such companions are the most dangerous for the planetary search because the path of the reference star will be only slightly curved. But those that are detected must be assigned very low weight until the orbit is well established. Thus a conservative approach is to begin with 15 to 20 stars in each star field. To enable coverage of maximum area, the detectors need the flexibility of vertical motion. We shall see that the accuracy in measuring the vertical position is not critical for the proposed detector scheme. Coarse adjustment and no further motion during observation should be possible. The detectors have finite vertical dimension so the stars must be selected on a noninterfering basis for both telescopes. When the detailed detector design is undertaken, a small excess of detectors should probably be provided to attain enough stars. One possibility is to install two detectors per rail and an optical flat at each end of the rails shown in figure 37. A laser interferometer would be needed at each end of each rail set plus one or two more interferometers to check the distance between the two optical flats. The extra complication of aligning and monitoring a second reference surface must be considered in such a choice.

ISI Detectors

Tracking– Figure 38 is a sketch of an individual detector or unit and an associated Cartesian coordinate system. The device is mounted on four preloaded, "open-type," linear ball bearings and is free to move in X on the two outer round ways over the entire field of view (1.75 m). The cylinder (15 cm long, 9 cm outside diameter) containing the optics and imaging tube is free to move on Y ways over a distance of 9 cm centered on the central way. Small movements in Z (~3 mm) are also possible for focusing purposes. The required motions in X, Y, and Z and a means for obtaining them are discussed below.

The central way shown in figure 38 is associated with the X drive mechanism. The drive should reach a constant velocity of ~0.7 cm/sec in 0.5 sec from rest with the capability for small corrections (1 mm or smaller) during an observation. The total slop in X should be less than 10 mm. Transients should be sufficiently damped within 1–2 sec of startup; additional transients introduced by the correcting mechanism must be kept small. Capability for fast return should be included to maximize observation time.

Three possibilities for the X drive will be discussed: (1) piezoelectric drive, (2) lead screw, and (3) servo-chain drive.

Piezoelectric drive: This drive makes use of a piezoelectric "inchworm" (Burleigh Instruments, East Rochester, New York) sketched in figure 52. This device uses a sequence of electrical pulses to effect a series of mechanical deformations of the quartz crystal (as shown in the figure). At the completion of one sequence, the device has taken one step along the rod. In the present application, a slotted cylindrical device would be used to grip the third way and translate the device in X.

The current specifications for this device are given in table 9. Current technology does not include a device with sufficient speed, overall length of movements, or proper geometry for the proposed application. Slow return slewing speeds would also limit observation time. However, the inherent simplicity and accuracy of the device make its use desirable, and a segmented device of many central segments should be investigated as a means of overcoming the above deficiencies.

Lead screw: A lead screw with either a double-nut preloaded ball nut or spring-loaded Teflon rings serves as another drive alternative. Rohlix drives also provide a similar function.

Servo-driven chain: A servo-driven chain running in a groove cut in the top of the central way and attached to the detector at its mass center is a third alternative. The chain is driven by one or possibly two dc servo motors and a linear encoder permits feedback position data. The cost of this system is expected to be less than the others described above.

The Y position is fixed during an observation and is approximately known for each star expected within the field. However, some correction in the expected Y position may be necessary during

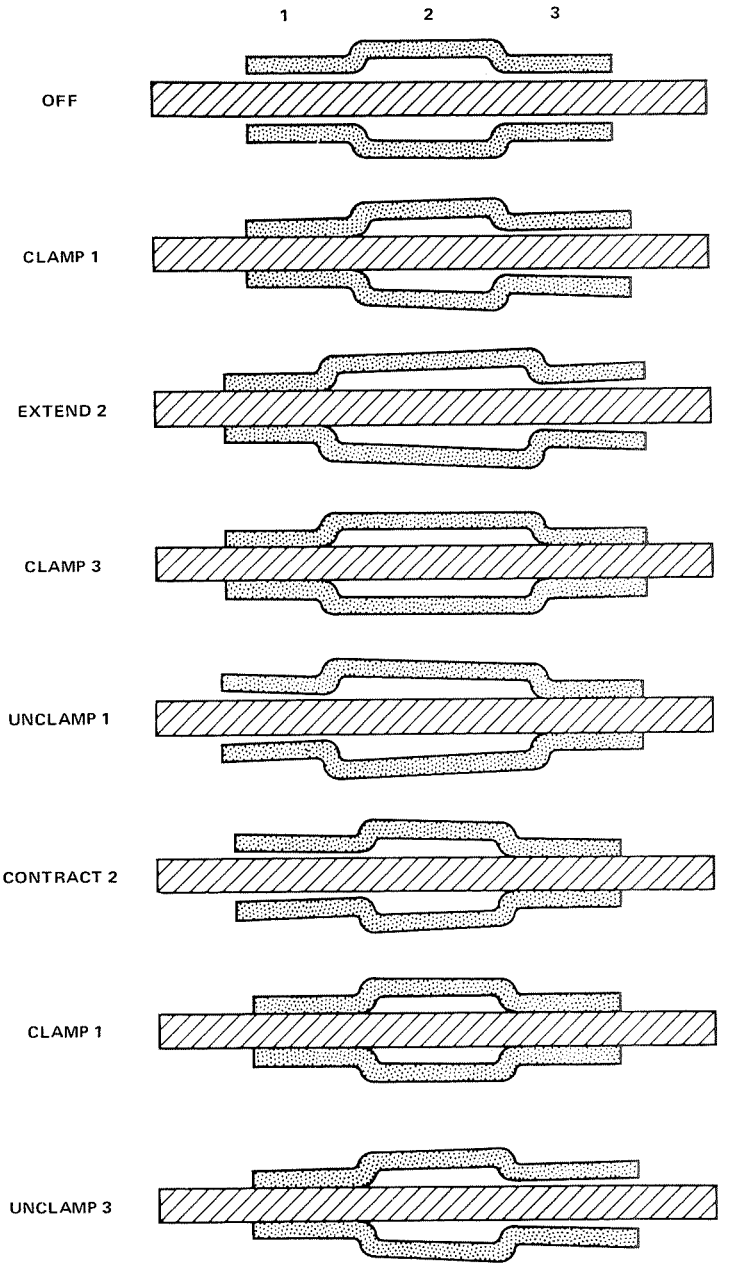


Figure 52.– *Schematic view of the operation of a piezoelectric drive mechanism (see text).*

TABLE 9.– SPECIFICATIONS OF INCHWORM

Travel	0–25 mm
Resolution	6 mm
Speed	3–20 mm/min
Lateral spindle displacement	0.5 mm (typical)
Maximum axial load	2 kg
Housing material	Chrome-plated stainless
Spindle and spindle tip material	Invar 36
Size	0.2 by 0.02 m (8.8 by 0.8 in.)

the initial observations of a given night. The Y motion of the optics will be produced by the inchworm device.

The focal plane is a doubly curved surface of 100-m radius of curvature. The required Z position of the optics, needed for focusing, is a known function. The motion may be provided by a linear actuator situated behind and driving the cylinder containing the optics/image tube along its axis. Driving the actuator by a dc motor/encoder/gear box combination will provide a 1.0-μm resolution. The cylinder would be supported on ways and would be spring-loaded to minimize backlash.

The X position of the optical head and its tilt about the Y axis must be determined with interferometric accuracy. The proposed laser system and flat reference mirror are shown in figure 38.

The flat mirror extends over the entire height of the detector region and serves as the fundamental reference relative to which measurements of X and tilt about the Y axis are made. The tilt measurements provide a small correction to X position measurements. It is necessary to know of the tilt around the Y axis because of the roughness of the ways – of the order of a few nanometers. These changes in the orientation of the optic axis of the collimator multiplied by the focal length F_2 become changes in the apparent X of the traveling microscope. The easiest of several ways to monitor this tilt is to

mount two cube carriers on each platform in the XZ plane and to run two laser interferometers on the platform, referring each to the reference flat.

Detector optics– The optical system within the detector is sketched schematically in figure 53 and orthographically in figure 54. The system is basically a wave-front folding interferometer in which the beams are modulated and mixed as described below.

Light from each of the primary mirrors is first passed through a chopper which is part of the star acquisition scheme discussed below. Lens L1 serves to collimate the two beams, after which they enter the Koester prism, K1, are reflected, and passed on to roof 1 or to the internal reflecting cube corner. Roof 1 serves as a piezoelectrically driven modulator. A phase difference of π radians exists between light entering prism K2 from the cube corner and light entering K2 from roof 1. Koester prism K2 contains a beamsplitter (mixer). The light beams now pass out of K2, still collimated, and enter roof 2, which redirects the beams. Roof 2 also has a diffraction grating in the path of the emerging beams which disperses the light. The dispersed light enters lens L2, which throws the pupil to infinity and forms an image of the star at distance F beyond L2. Here F is the optical path length from the mixer to lens L2.

Lenses L3 and L4 are at a distance $2F$ from L2; lenses L3 and L4 have focal lengths F/2 and F, respectively. Lenses L3 serve to image the star in the photocathode. Lenses L4 are central slabs cut from convergent lenses that serve both to define the passbands of blue, yellow, orange, and red light and to image the entrance window on the image tube photocathode. To avoid a dispersed image of that entrance window, a blazed grating identical in line spacing to the first grating is placed under lenses L4. It completes the recollimation of the beams that left the exit pupil at the beamsplitter. The image tube is at distance F from lenses L4.

Detector electromechanical functions– As outlined previously, an optical delay line is realized in one path of the detector interferometer in the form of a roof prism. It will be scanned electrically over an optical path difference of about 50 wavelengths in about 0.5 sec in a sawtooth pattern. The delay line will be moved at, say, 56 mm/sec. Unfortunately, at least for the total scan distance visualized above, the integral linearity error is apt to be ±5 percent. Repeatability can be quite good – less than 1 percent – but the

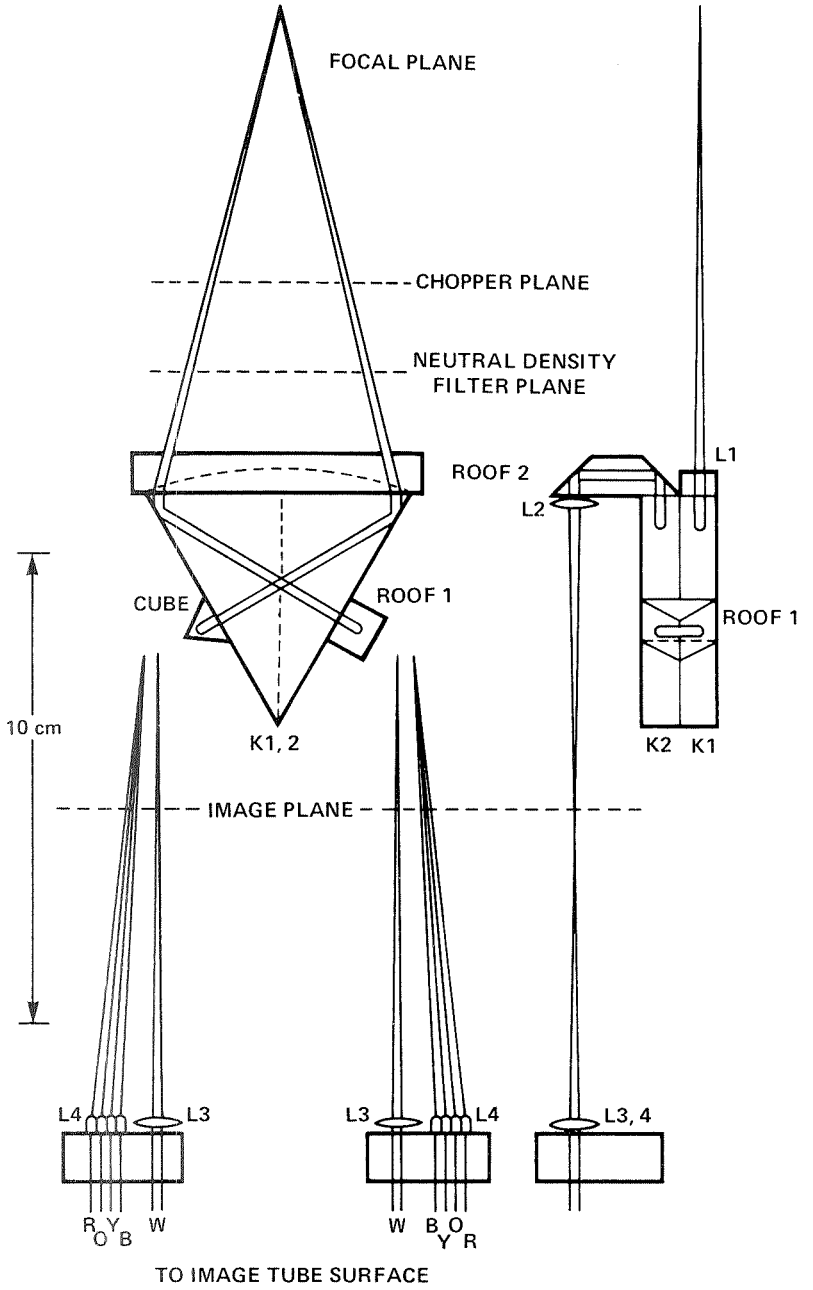


Figure 53.– *Detailed schematic of the optical system within a detector.*

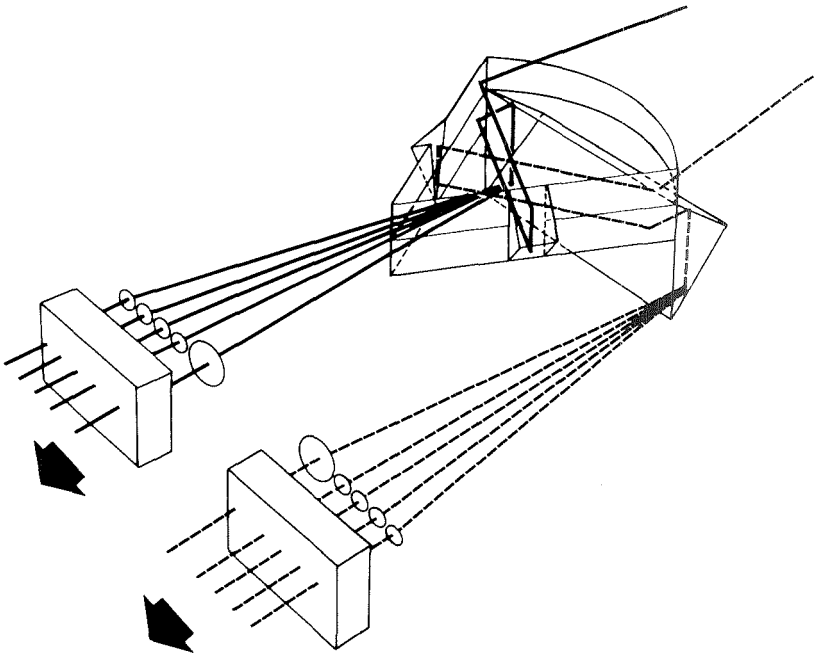


Figure 54.– *Orthographic view of a detector optical system.*

linearity in the voltage-to-displacement characteristic usually creates a problem. This can be easily corrected by storing the inverse nonlinearity in a read-only memory and performing the required transformation.

Figure 55 is a block diagram for a piezoelectric (PZT) delay line scanner. Upon receipt of a synchronizing pulse from the central data processor, the up-down counter registers another count. The stored count is transformed into a different data word by the read-only memory (ROM), thus producing the required signal for a position change. The ROM is followed by a digital-to-analog converter that provides a voltage drive for the high-voltage amplifier. The voltage amplifier produces an output voltage swing of about 1000 V for the PZT drive. The resultant voltage is read by a digital voltmeter for position confirmation. This voltage is retransformed by another ROM, in which the PZT nonlinearity is stored, and sent back to the central processing unit.

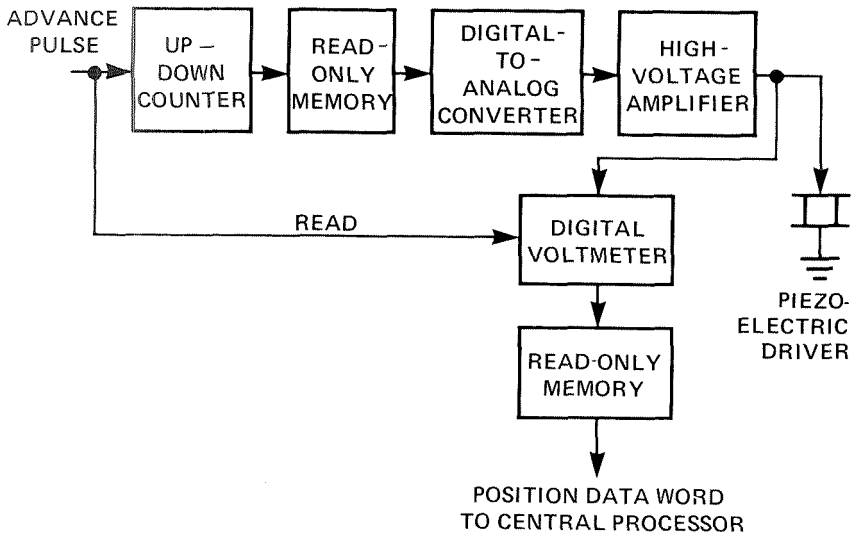


Figure 55.– *Schematic diagram of the drive mechanism for a piezoelectric transducer (PZT) which corrects for nonlinearities of the voltage to elongation characteristic (see text).*

The above process is repeated for each advance pulse and, when the up-down counter reaches full count, the mode is reversed. The delay line then sweeps in the reverse direction.

Detector configuration– Since space is at a premium in the tracking detectors, a primary consideration is size. Weight is also a prime factor since each detector must track its appointed star rather rapidly. Power consumption must be low for obvious logistical reasons. These parameters, when considered jointly, underscore the attractiveness of charge coupled device (CCD) imagers. Signal-to-noise considerations imply the necessity for image intensification (ICCD operation).

As mentioned previously, each image tube will process two white-light star images. In addition, each tube will have arrays to process four passbands centered, for example, at 0.45, 0.52, 0.59, and 0.66 μm. There are two arrays for each of these colors. The two arrays have the property that corresponding CCD elements are anti-correlated in output, a property we would like to exploit.

The overall configuration of the detector is depicted in figure 56. Light rays from the detector optics fall upon the photocathode, releasing photoelectrons on the other side. These electrons are accelerated toward the CCD arrays by a high voltage. When they reach the CCD arrays, they produce secondary electrons in the semiconductor substrate. This is a gain mechanism. Gains of several thousand have been reported at electron energies of about 10 keV or so (General Electric Company, internal report, 1976). Alternatively, the fast electrons impact on a phosphor on the output face of a sealed image tube and the subsequently emitted light is transported through a fiber-plate coupler to the CCD array.

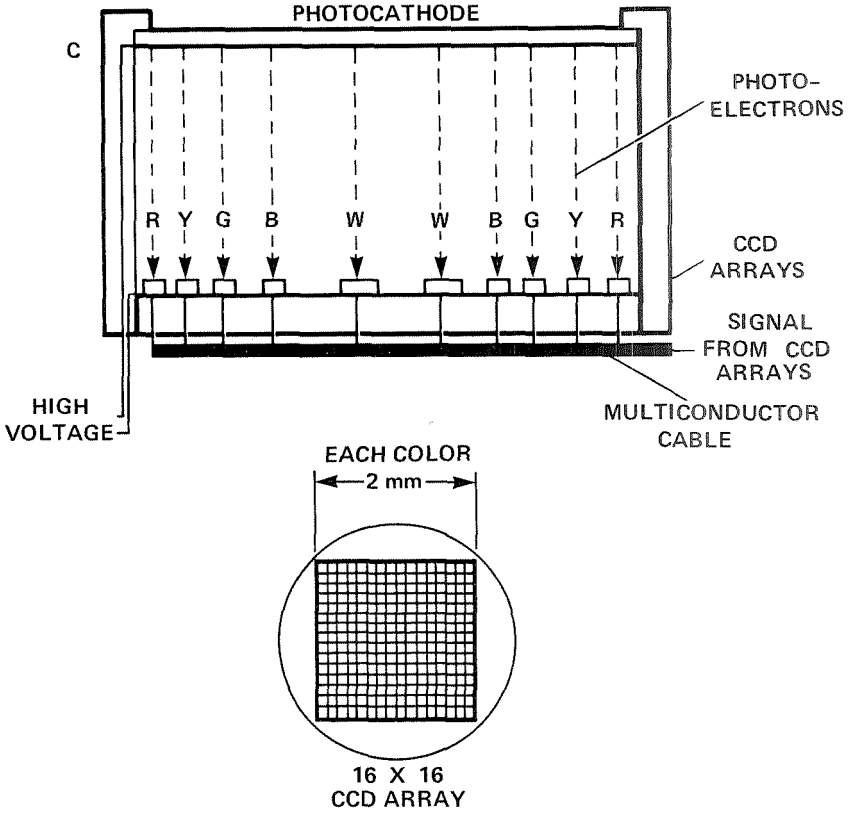


Figure 56.– *Detailed schematic of a photocathode/CCD system used in each detector.*

The sketch in figure 56 shows the arrangement of electron beams in the image tube. Each square array is a CCD imager assumed to be operating in the EBIC (electron-bombarded, induced-current) mode. An "on-chip" integrating sense amplifier will be provided for high signal/noise ratio. Information coming from the arrays emerges as discrete-time analog samples, 10 lines in parallel. Their further processing will be described subsequently.

Detector functions– As pointed out in a previous section, there is space for housekeeping functions in two compartments between the X ways. Above them is a slide that moves in the Y direction upon which the optical equipment is mounted. This section of the report outlines the electronics and logic functions that are physically distributed in the moving detector carriage shown in figure 57.

A receiver antenna is mounted on the compartment beneath the X ways. It feeds a demodulator, and the demodulator drives a demultiplexer. The demultiplexer derives the following commands in the form of data bit streams: (1) a move pulse to cause the delay line to increment by one-quarter wavelength, (2) an array read clock to shift information from the CCD arrays, and (3) a data stream to actuate the focus, or Z axis, controller.

The delay-line logic is performed in the same package as the data receiver, the objective being to remove heat-generating and space-consuming components from the vicinity of the optics. The high-voltage drive to the PZT actuator will be cabled up to the slide. The array read information will also be fed directly to the slide, along with low voltage for the integrated circuit logic package on the slide and high voltage for the accelerating portion of the image tube. Low voltage will be transmitted to the moving carriage along the X ways and will be picked up by wipers. This voltage will be regulated and also transformed to higher voltage (several thousand volts dc). Because wipers sliding on metallic ways generate noise spikes, care should be exercised in voltage regulation and in the suppression of noise spikes.

The instrument package aboard the Y-slide is a multifunctional one. It accepts the high-voltage drive and actuates the PZT element. The DVM and ROM part of the delay line actuator logic resides here, along with a buffer to store the position data. These data are fed into a multiplexer along with the array outputs after analog-to-digital

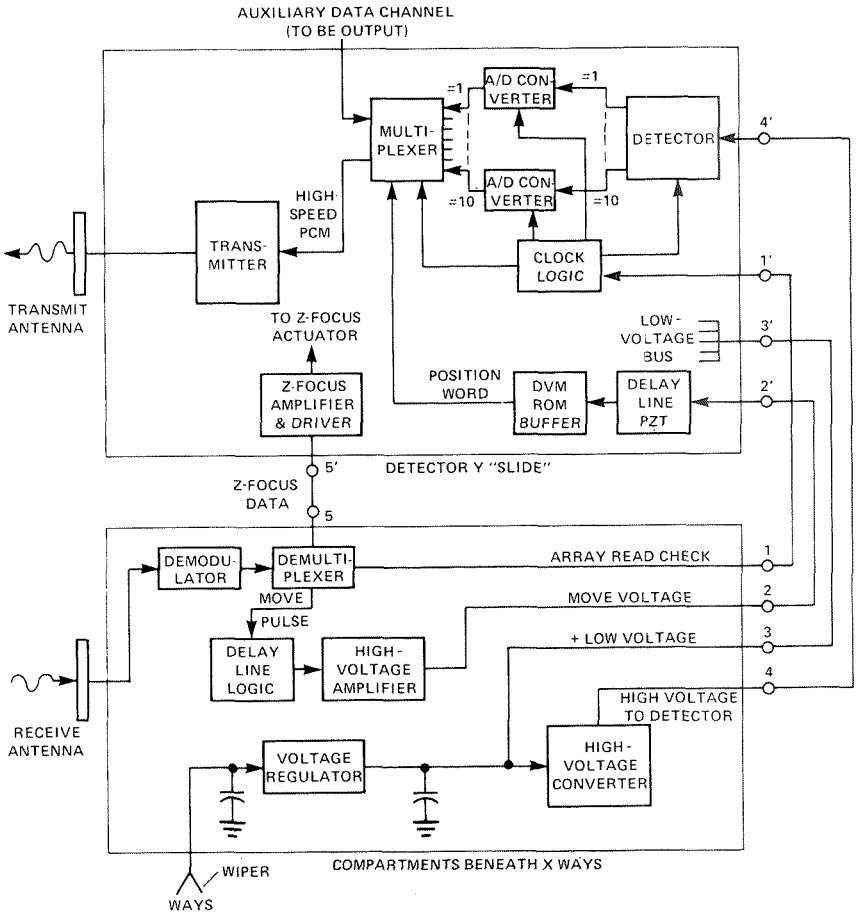


Figure 57.— *Schematic diagram of logic functions used in each detector (see text).*

conversion. These composite data are then modulated upon an rf carrier and fed to the transmit antenna affixed to the *Y*-slide.

Additional electronics functions housed in this portion include the clock logic to derive all on-board control signals plus the drive for the focus, or *Z*-tracking, controller. An auxiliary data channel is provided for any other functions desired.

Data Processing

Summary of data handling– Data routing, acquisition, fringes, and traveling detector positions and tilt are summarized below.

Data routing: A schematic functional diagram of data routing on board the moving detector is illustrated in figure 58.

Acquisition: In order to acquire its appointed star, each detector must initially be placed in an acquisition mode. Because its operation has been described elsewhere, it suffices here merely to mention that the data required for acquisition are contained in the bit stream from the white-light channel arrays. Hence, these data are routed to the central processor via the multiplexed data stream.

Fringes: As the delay line changes position, the temporal fringes in each beam of colored light make the image of the superimposed input windows appear to blink, depending on the color of the light, at about 100 Hz. Varying intensity is sampled at each photosite in the CCD detector array. The bit stream emanating from the detector array represents digitized versions (8 bits/sample) of the detector photosite intensity.

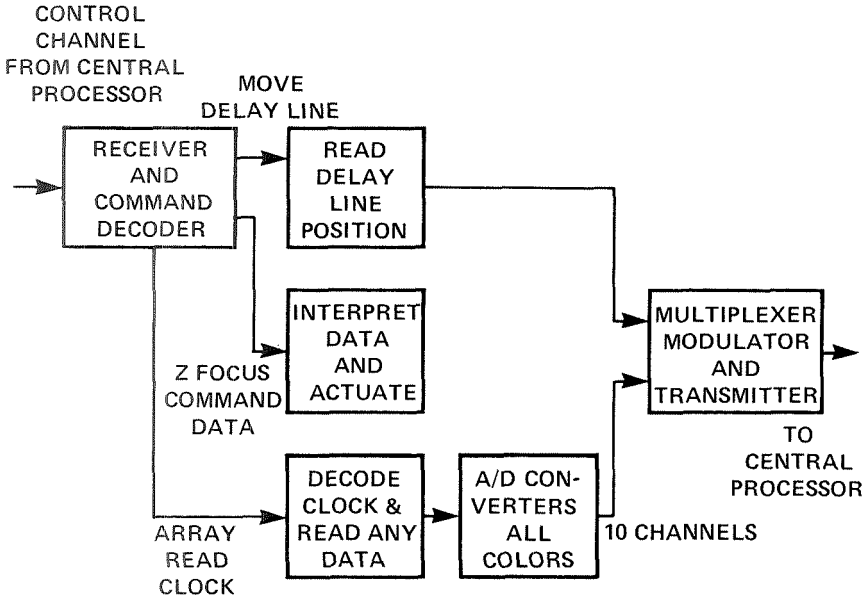


Figure 58.– *Schematic diagram of data routing within each detector.*

Traveling detector position and tilt: A laser interferometer means is provided for ascertaining the X position and tilt of each detector. Such a scheme is illustrated in figure 59. The interferometer consists of a Koester prism, as in the detector optics. Outputs X and Y are electrical signals complementary to one another. The succeeding electronics package removes a carrier wave from X and Y, counts low-frequency fringes relative to a calibrated origin – in a *bidirectional* sense – and reads each zero crossing to approximately one-twentieth of a wavelength[4] (Hewlett-Packard Company, application No. 197-1, 1975). Note that the bidirectional requirement is a consequence of the occasional retrograde X motion of the detector. The laser and the signal demodulators can all be arranged to be outside the vacuum.

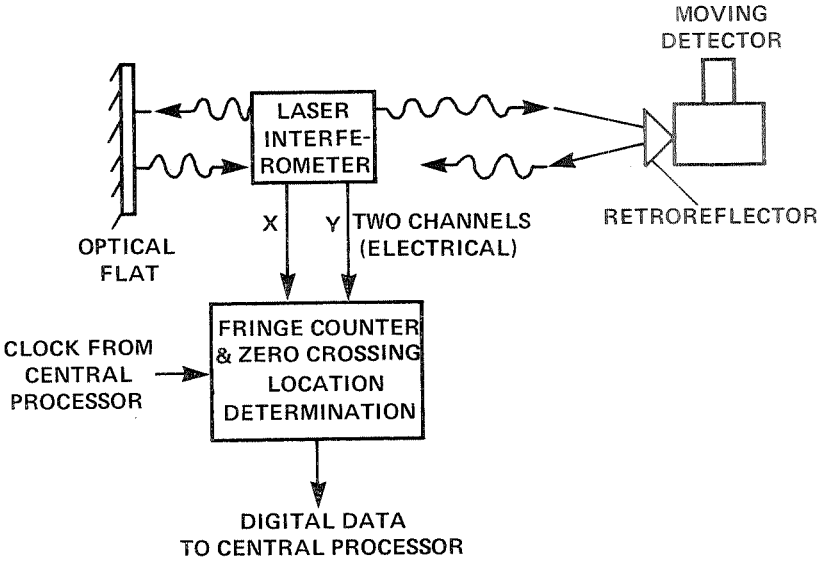


Figure 59.– *Schematic diagram of a laser ranging system to locate the X position of each moving detector.*

[4] A fringe-counting system is currently being manufactured by the Hewlett-Packard Company. Using Doppler shift techniques, it provides the precision required. Its main use, at present, is in machine tool positioning systems and step-and-repeat cameras for generation of integrated circuit masks.

Data reductions for star image position– The signal seen by a CCD element in one of the eight images of the entrance window is of the form

$$I = U(1 + V(t - t_0)\cos 2\pi(t - t_0)/\lambda)$$

where t is a relative path difference between the corresponding points in the two entrance windows that have been superimposed by the folding of the wave front about a line midway between the windows. The modulation depth V, which has a maximum $t = t_1$, corresponds to a visibility function in a Michelson wavelength spectrometer if the beamsplitter divides the energy evenly between the two arms. Otherwise, the modulation depth is proportional to the smaller fraction in a split of fractions f and $1 - f$. The dc level U contains no position information, but is of interest for photoelectric photometry of the star.

In space or on the Moon, we would scan the path difference and find the star position in the field of view of the microscope from t_0 as a result of a least-squares fit of $I(t)$ to the entire waveform. But the atmosphere moves the star position, as we saw in an earlier section, by an amount corresponding to $t = 10\lambda$ at frequencies up to 1 Hz. This implies that fringe speeds up to 20π waves/sec are common. That means that if we modulate t at 100 waves/sec, there will be some moments when the star position just keeps up with the intentional motion and successive samples of I will be the same because $t - t_0$ is slowly changing. And there will be other moments when the apparent modulation rate is doubled or more. Clearly, we need a scheme that pays attention to a less rapidly varying quantity than t_0. Such a quantity is the modulation depth $V(t - t_0)$. If $Q = \lambda/\Delta\lambda$ is a measure of the width of the spectral bandpass falling on one detector, then V falls appreciably below its maximum value outside a region of $t - t_0$ about Q waves wide. More precisely, the modulation depth V is the Fourier transform of the effective energy distribution incident on the detector, including all instrumental effects such as wavelength-dependent quantum efficiency and so on. We cannot avoid having to sample I at least 200 times per second, which means few photons and high data rates from 256 detectors per exit pupil and 8 exit pupils. But we can hope to sum at least Q of those samples in some manner that reveals an average of V in a

"small" region of its variation. That will decrease the data rate by a factor of about Q.

There are several possible schemes to detect the modulation depth V. Basically, we need to detect the fluctuating part of I, rectify, and sum. The output will include a noise component, which tends to make it difficult to locate the path difference t where V has its maximum. But, in the case of quantum noise, it can be shown that the noise is correlated with V, so it may be expected that the method will work even if there are few counts per Q samples per exit pupil array. For example, for a fifteenth-magnitude star in a 0.05-μm passband centered at 0.5 μm, there are about 500 electrons per second per aperture of πm^2, assuming 15-percent quantum efficiency and 50-percent transmission after reflection and grating losses. Combining the two images of the input window in that color, we find 50 counts per 1/20 sec. If the atmospheric phase fluctuations do not exceed $Q\lambda$ nor appreciably decrease the depth of modulation, then the formal uncertainty in star position after the 300-sec integration would be $Q\lambda/(BN^{1/2}) = 50$ μarcsec, where $N = 1.5 \times 10^5$ is the total number of electron counts. But we saw earlier that the atmospheric phase fluctuations are indeed bounded by about $Q\lambda$. And if the modulation scheme is as rapid as called for in the previous paragraph, it is clear that appreciable decreases in the modulation depth will be occurring less than a third of the time. Thus we satisfy the formal requirements at the fifteenth magnitude for small to negligible uncertainty in the star position at the 100-μarcsec level due to quantum statistics.

In discussing possible schemes to detect the modulation depth V, let us first understand that all four passbands can be studied simultaneously, despite the fact that the roof speed is not ideal for any passband. Let the effective wavelengths be $\lambda = 0.45, 0.52, 0.59$, and 0.66 μm as listed in table 10. Suppose the roof speed is defined by a linear sweep of 14 μm in 0.5 sec. Then the rate of optical path change is 56 μm/sec, or 100 λ/sec at the intermediate wavelength of 0.56 μm/sec. Then at $\lambda = 0.45$ μm the fringes go by 24 percent faster. Thus, if a sample No. 1 were centered at 0 phase, sample No. 3 would be at 180° for $\lambda = 0.56$ μm but at 224° for $\lambda = 0.45$ μm. Now the cosine of 224° is −0.72 instead of −1.00, so if we subtracted sample No. 3 from sample No. 1 and kept only the absolute value as a measure of the modulation depth, we would have an estimate that

TABLE 10.– MODULATION DEPTHS AT FOUR WAVELENGTHS NEAR 0.56 μm WHERE SAMPLING IS AT A RATE OF 4 SAMPLES PER CYCLE AT 0.56 μm AND ALTERNATE SAMPLES ARE SUBTRACTED

Wavelength, μm	$v = (1 - \cos \pi(0.56))$ $\lambda/2$	$v = \sin(0.56\ \pi/4\lambda)$ $(0.56\ \pi/4\lambda)$
0.45	0.860	0.804
.52	.985	.900
.59	.994	.884
.66	.944	.815

is 86 percent of the maximum estimate (column 2 of table 10). But the fringes are moving. If we sample the light output p times per cycle, then the blurring decreases the apparent depth of modulation by a factor of $\sin(\pi/p)/(\pi/p)$. The apparent V for this sampling scheme is column 3 of table 10. The passbands closest to 0.56 μm have the deepest modulation, but all have modulation greater than 80 percent when the star is not appreciably resolved and the fringe speed due to the atmosphere is zero. Such a sampling scheme is motivated by the fact that, if a sinusoidal wave is sampled four times per cycle, then the least-mean-square fitting procedure involves sines and cosines that are all ±1 or 0. Thus, digital multiplication can be replaced by the much more rapid addition or subtraction. Indeed, if the four samples I_j in one cycle are to be fitted to

$$U + a \cdot \cos\theta + b \cdot \sin\theta$$

we find, in the absence of blurring,

$$a = (I_1 - I_3)/2$$

$$b = (I_2 - I_4)/2$$

$$U = (I_1 + I_2 + I_3 + I_4)/4$$

$$V = (a^2 + b^2)^{1/2}/U$$

then if we take $UV = |a| + |b|$, we approximate the hypotenuse of a right triangle by the sum of its two sides. For random fringe position,

this overestimates UV by $4/\pi = 1.27$ and gives a special role to the phases that are close to odd multiples of $\pi/4$ where the error in modulation depth is largest, 41 percent.

In a second possible sampling scheme, suppose that we have time after an analog-to-digital conversion of each CCD output to form I^2 by digital multiplication. Successive values of I^2 for a given detector can be represented by

$$|U + V \exp(i\phi)|^2 = U^2 + V^2 + 2UV \cos \phi$$

Ignoring the atmosphere, we see that for the passbands of table 10, the average of I^2 for four samples will be significantly different from $U^2 + V^2$ because the average of $\cos \phi$ will be quite far from zero. However, it is easy to select a time interval about Q half-cycles long at $\lambda = 0.56$ μm for which the average of $\cos \phi$ is less than 1 percent for each of the four passbands. Since such an interval is adequately small for a measurement of the center of V, we can estimate V^2 from $\langle I^2 \rangle - \langle I \rangle^2$ with negligible systematic error in any passband. In fact, because of the slow drift of $\cos \phi$ in that averaging period, the sampling frequency could be decreased from 400 to 200 Hz, which is the Nyquist frequency for $\lambda = 0.56$ μm. The danger of sampling a pure sinusoid at exactly the Nyquist frequency is that, instead of sampling maximum-minimum-maximum and so on, the sampling might be mean-mean-mean and so on. No such danger exists for passbands not containing $\lambda = 0.56$ μm and is irrelevant for a long-term average in the presence of atmsopheric perturbations larger than $\lambda/2$.

In a third possible sampling scheme, we recall that, for corresponding detectors in the exit pupils of a given passband, the outputs are proportional to, say, $U(1 + V \cos \phi)$ and $RU(1 - V \cos \phi)$, where R is a factor for relative sensitivity of the two detectors. Then

$$U = (I_1 + I_2/R)/2$$

$$UV \cos \phi = (I_1 - I_2/R)/2$$

It is clear that if the relative sensitivity factor R were close enough to unity for every detector pair, we could use the complementary outputs for the separation of the dc level and the modulation by the

simple process of digital sums and differences of simultaneous outputs from corresponding detectors (which is not trivial from a programming point of view because the exit pupils are mirror images, but the CCD arrays are not apt to be read out in that order unless we require it at manufacture – a stipulation that would increase cost). Experience with diode arrays at the University of Arizona suggests that deviations of R from unity are apt to be about 5 percent. This suggests use of the positive definite statistic:

$$|I_1 - I_2| = U(|1 - R| + V(1 + R)|\cos \phi|)$$

which increases with UV above a constant but small bias of $U|1 - R|$. Such a statistic is suitable for finding the position of the maximum value of V with minimum complexity provided each $R \cong 1$.

To compare the first and second sampling schemes, it is enough to show that present hardware is capable of meeting the timing requirements of the second scheme so that the shortcuts of the first scheme are not necessary. Suppose 200-Hz (Nyquist) sampling, so that there are 5 msec per cycle in a 500-msec sweep through 14 μm of roof motion. One possibility among many is that a microprocessor is provided for each 128 elements of a photoarray. With 256 detectors per exit pupil and 10 such arrays, this calls for 20 microprocessors. The time allotted for each detector, then, is 39 μsec. In 39 μsec we must do an 8-bit analog-to-digital conversion, add I_J to the appropriate $\sum_{j=1}^{10} I_j$ register (12 bits), form I_j^2, and add it to the appropriate $\sum_{j=1}^{10} I_j^2$ register (20 bits). That time needs to be allocated so that after 10 cycles the results in the summing registers can be shifted out to buffers for transmission to a general purpose minicomputer. The transmission rate is then, ignoring imaging array used for acquisition, 16 arrays by 128 detectors by 32 bits/50 msec = 1.31 M bits/sec for each detector. A digital multiplier capable of forming a 16-bit product in 13 μsec is manufactured by American Micro Devices, among others. An 8-bit analog-to-digital conversion in 13 μsec is reasonable, so that leaves 13 μsec for two additions, some counter-increments, and logic checks. If this 39 μsec were somehow not adequate, we could double the number of microprocessors unless heat dissipation and volume prohibited.

In the above example, the next processing task would be to calculate the total variance $S = \Sigma I^2 - (\Sigma I)^2/N$ for each detector. There would be some advantage of calculating it before transmission because it is not apt to be more than a 12-bit number. Total variance and ΣI could also be averaged with the corresponding detector in the conjugate exit pupil and further decrease the bit rate by 1/2. Then the set of numbers $S(t - t_0)$ for each detector can be combined for an estimate of t_0; the example has 10 numbers per sweep for four colors and 256 detectors per color, requiring only 10,240 numbers in core. It seems probable that the fastest way to obtain an estimate of t_0 involves a fast Fourier transform of S, because the lowest Fourier component could be slightly corrected in phase for instrumental effects in the optics before averaging all the detectors in an exit pupil. The correction in phase would require only four multiplications per detector after looking up at most 2048 numbers in a file. The phase corrections should markedly decrease blurring of the t_0 distribution. After averaging the t_0 values over the exit pupil, we are left with four values of t_0 and Δt_0, one for each color. For photometry we would also like to record the average over the pupil of ΣI, where correction is included for relative sensitivities of the detectors, if needed. The large data rates per star suggest that each transmitter be followed by an arithmetic and memory unit per star that is slaved to the central processor. Data rates are low when t_0 becomes available. The t_0 values are to be monitored by the central processor to ensure that each traveling microscope follows its star successfully.

ISI BUILDING

Long-term stability is essential for both the site and its supporting structure. To achieve this stability, it will be necessary to obtain a firm bedrock site condition along with a massive and continuous reinforced concrete foundation.

Foundation Construction

Many alternative schemes are possible and the following basic structure is given to provide at least one workable method for feasibility and cost analyses. There are three basic types of foundation elements:

1. Heavy reflector footings; 1 m thick and heavily reinforced to ensure stability
2. Vacuum tube footings; 0.5 m thick, moderate reinforcement
3. Intermediate working area slabs, 0.3 m thick, moderate reinforcement

Superstructure

After the foundation is completed and the walls are poured, the vacuum tubes and utility services are placed, the support structures for the heavy mirrors and reflectors are installed, and finally the optics are placed. The precast roof slabs are then installed, followed by waterproofing and insulation, for example, a 1-ft thickness of pumice material. Final earth embankment, grading, and landscaping should then be completed.

The surrounding site grounds should also be covered with 0.15–0.30 m (6–12 in.) of pumice or other suitable, vesicular rock to insulate the entire installation and to provide the least absorption contrast with the surrounding environment.

Vacuum System

To minimize refraction changes and resulting distortion along the horizontal light path, it is necessary to enclose this path in a vacuum. The proposed design encloses the primary mirrors, secondary mirrors, and detectors in an evacuated tube, leaving the tracking flat mirrors outside the vacuum system. The entire vacuum system will be evacuated to a pressure of 1 Torr (1 mm Hg) which is relatively easy to reach with existing equipment. The temperature and vacuum systems must be able to keep the vertical temperature variation within the tube to less than 1°C.

SITE SELECTION AND DEVELOPMENT

For any astronomical observatory to be successful, an optically sound telescope must be located at a site that minimizes the negative effects of atmosphere and weather on individual observations as well as on the observation program.

Site-Selection Criteria

High percentage of nighttime clear weather– To maximize the number of astrometric observations, the nighttime sky at the site should be as cloud-free as possible. A good site would have skies with less than 15-percent cloud cover for 75 percent of the potential observing period.

Maximum number of hours of astronomical darkness– The prospective site should have long, continuous hours of darkness spread rather uniformly over the entire year. Therefore, if only one site is to be selected, it should lie within latitudes 40°S to 40°N.

Dark sky– Lights from surrounding urban or industrial areas tend to increase the sky brightness to the detriment of good astrometrical observations. Thus, the site should be located away from current or potential land uses which result in the generation of high-intensity surface lighting.

High transparency– The sky at the site should be free from natural and man-made atmospheric pollution. Man-made pollutants could be in the form of smog, haze, smoke, aerosols, etc. Natural pollutants could be in the form of pollen, dust, ash, sand, steam, water vapor, etc.

Minimum optical turbulence (good seeing)– The site selected should be located in a region with a low level of wind activity during observation periods. Winds that do exist at the site should flow over the site in a smooth and laminar manner. A high, flat plateau may be far superior to a mountain. Site study should include both astrometric study of photographic star trails and correlation studies of arrays of sensitive barometers.

Control of future development of the site and its surroundings to ensure continuation of optimum conditions– Although all the

conditions listed above may be satisfied at the time of site selection, it is important that consideration be given to the possibility of future developments which could result in a deterioration of optimum site conditions. An example of an observatory where observing conditions gradually deteriorated due to surrounding development was the Lick Observatory located on Mt. Hamilton in California.

Accessibility– The site should be accessible to the personnel who would use the facility as well as to site-construction workers. Again, high mountaintops may be undesirable. We also will assume that the site selected should be located in the United States.

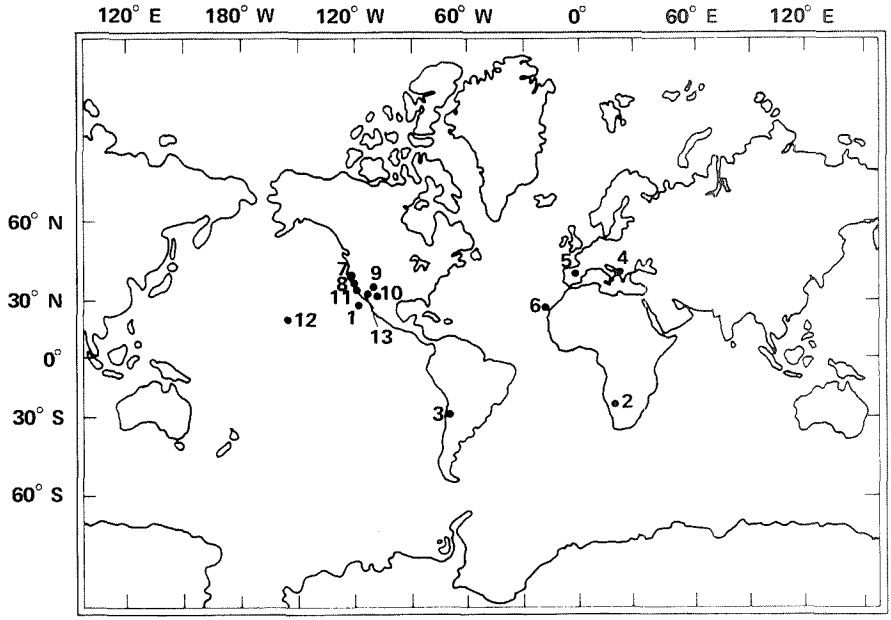
Ground stability– The site selected should be reasonably free from earthquake and volcanic activity as well as from microseismic activity.

Severe weather– The site should be free from severe weather phenomena such as hurricanes, tornados, typhoons, and duststorms.

Comparison of Possible Sites

Many site surveys have been taken around the world in the last decade. Climatological and seeing conditions at many sites have been extensively surveyed to determine their quality. The seeing was generally measured by use of the polar star trail method developed by Walker. Representative subject sites are shown in a worldwide map (fig. 60).

The high desert country of southwest United States is promising because of the prevalence of clear days and nights, and because of the high elevation with large flat areas. We suggest that the site should be located on a flat-topped mesa flanked by talus slopes for a major fraction of its height to provide a low turbulence atmosphere. We considered the southwestern corner of Arizona in a band with its center at Yuma, running to the north and east of Yuma for a distance of 30 to 70 miles. This area has the highest percentage of clear weather in the United States. In this region, the Kofa Mountains and the Castle Dome Mountains have high plateaus and low potential for future development. A significant portion of this region is contained within the Kofa Game Range, which is under the control of the Federal Bureau of Fish and Wildlife Management and probably would not be a potential site. All lands to the east of the game range



EXISTING OR POTENTIAL OBSERVATORY SITES

1. GUADALOUPE ISLAND
2. GAMSBER – S.W. AFRICA
3. LA SILLA – TOLOLO – CHILE
4. PARNON MTNS. – GREECE
5. SIERRA DE LOS FILABRES – SPAIN
6. FUENTE NUEVA (LA PALMA)
7. LICK OBSERVATORY, CA
8. JUNIPERO SERRA MTN., CA
9. USNO, FLAGSTAFF, AZ
10. KITT PEAK, TUCSON, AZ
11. MT. PALOMAR, CA
12. MAUNA KEA, HAWAII
13. PROPOSED SITE

Figure 60.– *Mercator projection showing the location of existing or possible sites for the Orion ISI. The proposed site (13) is at Black Mesa in Arizona (see text).*

are at altitudes too low for consideration. However, a mesa directly to the north of the game range looks promising. In particular, Black Mesa has the following characteristics: (1) an elevation of about 1100 m; (2) a flat portion in excess of 80 acres (100,000 m^2); and (3) it is south of an undeveloped road extending southerly from the main highway to within 1-1/2 miles of the easterly summit and within 1 mile of the westerly summit.

Site Development Considerations

Development of a site requires the following information: geological data including microenvironment or local lithospheric environment; meteorological or atmospheric conditions; and botanic or local ecology. Data from this investigation will be helpful in the final design stages and during operation.

Geological– The type and distribution of geologic materials must be determined in the near vicinity of the interferometer installation. Physical properties of the various materials in both the surficial unconsolidated and uppermost consolidated bedrock must be determined to minimize the natural microseismic motion of the foundation and to maximize the economical use of geologic materials. Rock properties to be determined include bulk density, porosity, saturation, bearing strength, composition, homogeneity, microseismic background, thermal coefficients, sonic velocities, groundwater table, transmissivity, and expansivity. An exploratory drilling program should be undertaken to obtain the necessary geological data.

A seismometer should be installed at the site for complete evaluation of astrometric data after observation begins. If the seismometer were operated for a year prior to the final design stages, it would be an excellent source of seismic data that are essential for detailed structural design of the optical element foundations.

Ambient Earth vibrations, or microseismic motion, are ever present in all natural and man-made structures. The amplitude and frequency of these vibrations vary from locality to locality, depending on numerous variables and natural phenomena. The ground motion can conveniently be recognized and discussed as short-period ($>$0.1 Hz), storm (0.1–0.01 Hz), and long-period ($<$0.01 Hz) microseisms. Microseismic data are shown in figure 61. Microseismic motion is least well known in the short-period and very long-period range because of the inability of existing seismographs to accurately record both high and low frequencies.

Ground motion arises from numerous natural processes. Causes that have been defined seem to have rather definite characteristic frequencies. The most pronounced type of motion is storm microseisms with peak amplitudes of 30–50 μm at 6–8-sec periods. These originate predominately from storms and atmospheric turbulence

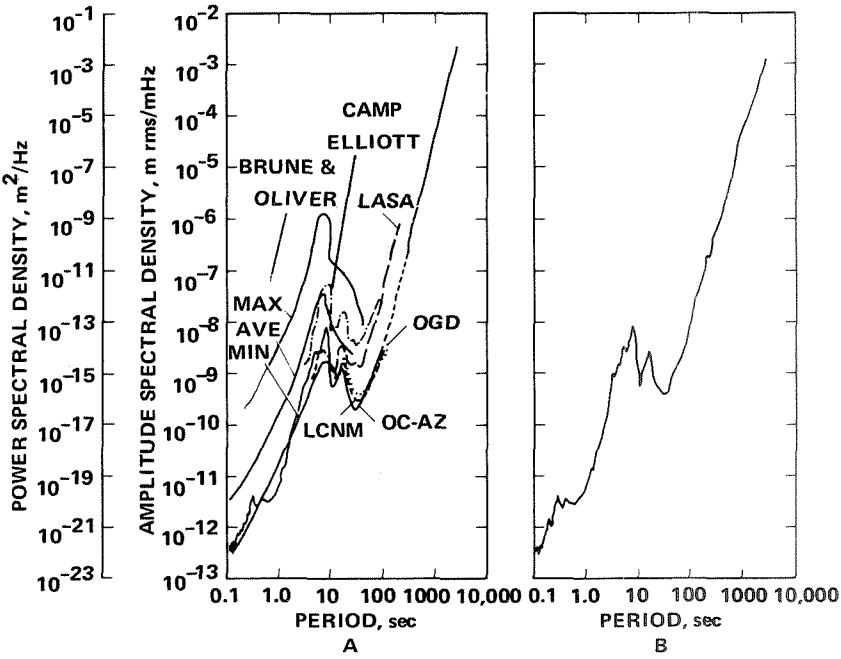


Figure 61.— *Data showing the amplitude spectral density (and power spectral density); spectral densities are shown as a function of motion period. (a) Vertical Earth surface motion. (b) Horizontal Earth surface motion.*

over ocean bodies, nonlinear interaction of ocean waves traveling in oblique directions, and nonlinear interactions of incident and reflected ocean waves near coastlines.

Long-period microseisms have been attributed to motion of ocean swells, direct transfer of ocean wave energy to sea bottoms, and, in some instances, to various meteorological phenomena.

Short-period microseisms are generated continuously by surf, wind, microbarometric oscillations, atmospheric fronts, thunderstorms, volcanic tremors, turbulent flow in rivers, water body noise (where the frequency is a function of water depth), Earth noise (tidal strain, mass movement, expansion, etc.), and cultural noise.

Variations in microseismic motion are seasonally, daily, and intermittently observed. The seasonal amplitudes are about 7 times greater during winter than summer at the 8-sec peak and are about

3.5 times greater at the 16-sec frequency. For the northern hemisphere, an amplitude increase begins in August and continues through winter. Daily variations are an exponential function of wind velocities above 4 m/sec (9 mph) and local barometric fluctuations. Because of the ocean wave contribution to microseisms (6–8 sec peak), ambient Earth motion diminishes approximately as an exponential decrease in amplitude as distance from the ocean is increased up to about 350 km. Cultural Earth motion due to such sources as cities, motor vehicles, railroads, aircraft, and dams may be significant locally up to 50 km.

Meteorological– Meteorological data should be obtained to provide frequency distributions of precipitation, wind velocity, and direction. Temperature and barometric data are required to ensure adequate thermal stability in the installation and also to provide the correct response time in the mercury base-level system.

Like the seismometer, permanent installation of atmospheric measuring devices should be an early project. This information would be computerized to provide automatic thermal control for the installation.

Botanical– A survey and mapping of the vegetation within 2 km of the site is necessary to ensure a minimum disturbance in natural surface conditions of the installation area. Although vesicular rock is the proposed ground material at the site, a transition must be created which will minimize thermal and meteorological differences between barren and vegetated areas. In addition, a program of vegetation removal, modification, or replacement must be formulated which will minimize the effect of the installation on the local ecology and natural beauty.

OPTIONS AND ADDITIONAL CONSIDERATIONS

Optics and Mirror Design Options

To achieve less error in correction for atmospheric effects, it would be desirable to increase the length of the tunnels in the telescopes to about twice the adopted 100-m length. All other components would remain the same size with suitable optical corrections to

the contours of the reflecting lens to adjust to the doubled length. The field of view would be reduced to half the diameter, but at the same time the errors introduced by the atmosphere would also be halved, provided the correlation length for atmosphere disturbance (cf. appendix A) is less than 50 m. Studies of that length for various sites and conditions will be needed first.

To reduce the cost of the telescope, all components can be made half size. This would reduce the total cost by one-half or more, possibly to as much as a third. The result would be a much less powerful system for determining star positions. Not only would the field of view be halved, but the light-gathering power would be reduced to one-fourth that for the proposed system.

Mirror Mounting Options

There are two basic classes of mirror mountings: active and passive. Active mirror mounts are required for large mirrors to minimize the effects of distortion due to gravity, temperature, wind pressure, and other environmental forces. Passive mirror mounts provide support only at the outer edges and so are used only for lightweight mirrors or when distortion is not a serious problem.

In many existing telescopes, the lever system has been used to support mirrors. This has been the practice for about 75 years and has been used with good success. The idea of the lever system is to provide a counteracting force to that of gravity at selected points on the rear surface of the mirror. The reflecting mirror lens of the 200-inch Hale telescope has 36 such support points. Furthermore, the lever systems are placed in weight-reducing recesses behind the mirror surface. The lever system is based on a compound mechanism. An outstanding advantage of the lever system is that it gives a reaction exactly proportional to the tilt angle of the mirror. Also, the system is essentially static and requires no adjustment under operating conditions. It suffers the disadvantages of friction and added weight.

In large mirrors, gravity effects result in distortions; therefore, recesses or voids can be formed in the rear of the mirror body to reduce mirror weight. Another technique would be to break up the

single large mirror into two or more smaller surfaces that provide sufficient reflection for the task at hand.

Air bags have the advantage of distributing support over larger areas. Another technique is to form a sealed cavity behind the mirror and support it uniformly by pressure from behind or, in the reverse direction, by a vacuum (assuming the mirrors are not in a vacuum chamber).

A more recent approach is that of a completely active mirror constructed so as to require continuous monitoring of its surface by means of forces provided by mechanical, hydraulic, or pressure techniques. Such a mirror surface would be quite thin and thus incapable of any self-retention of an accurate surface contour.

Counterbalancing of Tracking Flats

The large size of the four tracking mirrors presents a formidable problem in positioning and tracking. If the axis of rotation is chosen to pass through the center of gravity of the tracking mirror unit, then there is no need for counterbalancing forces. However, the point at which any given light ray impinges will then move as the tracking mirror moves and any surface imperfections will hinder the detecting process. Thus, the axis of rotation was located on the surface of the tracking flat mirror, resulting in a variable unbalanced moment. The magnitude of this moment is 2800 N·m at 30° tilt angle and 5300 N·m at 70° tilt angle. There are three means of overcoming this imbalance: (1) provide a motor of sufficient torque, (2) install a counterweight on an arm, and (3) attach a spring system to provide a counterbalancing moment. The choice made in this project was to select a suitably large hydraulic torque actuator.

ISI as a Geophysical Instrument

The proposed imaging interferometer can be used as a seismic instrument by analyzing the data differently. Astrometry studies differences in star coordinates. Geophysical phenomena are studied by analyzing coordinate sums – the average position of a star field is

sensitive to the rotation rate of Earth and to tilts of the interferometer. As a tiltmeter, the ISI would be limited by the recording sensitivity of the detector system and, on the high-frequency side, by the natural frequency of the geologic material upon which it is built or on the frequency of the designed foundation.

We did no detailed study on the use of the ISI in this mode, but a few remarks can be made. Current studies of the rotation of Earth would benefit from more precise transit circle devices, and polar wobble studies would benefit from better declination measurements. Both needs are met by the ISI. As a catalog of star positions for the ISI observation zone is gradually improved, it will become possible to tie together all the observations during a night to study small tilts in the interferometers, both cyclic tilts and tilts of longer trends. Such tilts result, respectively, from tides and lithospheric streams.

SUMMARY

The current level of precision (and accuracy) in astrometric observations is significantly poorer than can ultimately be obtained by a ground-based telescope. The ultimate limitation of the precision of ground-based astrometric observations of the type considered here derives from effects due to Earth's atmosphere. For systems that are small to moderate in size ($\lesssim$100 m), this atmospheric limitation is at the level of $>10^{-4}$ arcsec, a factor of 30 below current precision levels. Whether larger (1 km) systems such as that under consideration by Currie and co-workers can further reduce the detrimental effects of Earth's atmosphere is an open question. This factor of 30 can be achieved by utilization of modern technology in conjunction with well-established concepts of telescope optics. The Orion ISI is a novel and complex device that employs state-of-the-art technology and is theoretically capable of attaining 10^{-4}-arcsec/yr accuracy in relative astrometric observations.

As with any complex instrument, the Orion ISI has both weaknesses and strengths. However, in the admittedly biased opinion of the "Orion people," the latter far outweigh the former. The principal limitations or weaknesses of the ISI are:

1. Complexity
2. Restriction to relatively bright ($m_v \gtrsim +15$) stars
3. Restriction to stars that are more than 5–10 pc from the Sun.

It is difficult to assess the extent to which the intrinsic complexity of the ISI is a weakness of the system. It must be stressed that the Orion ISI concept is the result of only a few weeks of work, and that quite often the first attempt at design of a novel instrument is much more complex than it need be. Subsequent analysis of the Orion ISI concept may reveal that the essential aspects of the concept can be captured and recast in a less complex (hence more reliable and less expensive) form. The second limitation listed above stems most directly from the requirements for very short ($\lesssim$0.1 sec) integration times to minimize atmospheric effects. This limitation is not too severe. The third limitation, the "far-sightedness" of Orion, derives from the fact that the ISI requires fringes to obtain accurate position measures. If a star is resolved by the ISI, it will not produce such fringes. A typical stellar diameter is $\lesssim 10^9$ m (for spectral types G-M) and the resolution of the ISI ($= \lambda/2B = 5\times10^{-7}$ m/ 2×10^{2} m $= 2.5\times10^{-9}$ rad) is sufficient to resolve 10^9 m at a distance of $\sim$10 pc (3×10^{17} m). In terms of a comprehensive search, this limitation may not be too severe because the ISI could search for planets around stars out to a distance of 30–50 pc, and the number of stars in the volume of space bounded on the outside by a distance of, say, 40 pc from the Sun and on the inside by 10 pc from the Sun is $\sim 4^3 = 64$ times the number of stars within 10 pc of the Sun.

The principal assets or strengths of the Orion ISI are:

1. Deep space coverage
2. Stable structure
3. Freedom from optical aberrations
4. Large observing field
5. Minimization of atmospheric effects
6. Fully automated operation
7. Real-time data reduction
8. Ultrahigh precision

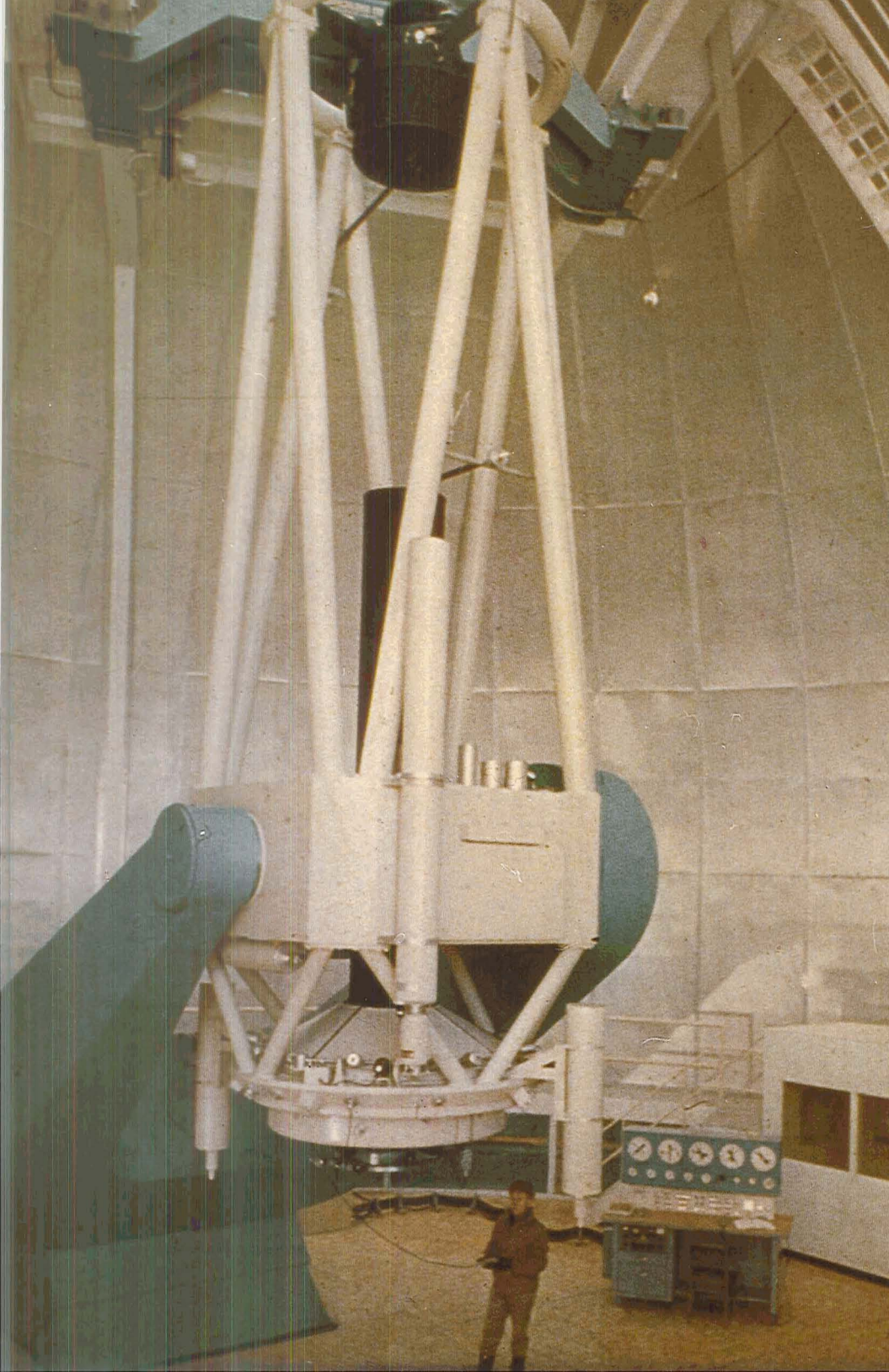
Asset (1), the ability of the ISI to study stars as far from the Sun as 30–50 pc, is equivalent to saying that the ISI can search a sufficiently large and homogeneous sample of stars to provide statistics on the nature of planetary systems. The highly stable structure (not

generally a characteristic for telescopes at this level of stability) means that there are a minimum number of moving key optical components (tracking flats) that are subject to varying gravitational stresses, and also that the thermal environment of the entire system is controlled. The ISI is, in principle, free of those optical aberrations (e.g., coma and astigmatism) that are major sources of astrometric error. The ISI instantaneously images a 1-square-degree field of view (2 square degrees during a 5-min observation). This field of view provides typically 10–20 *simultaneously* measured reference stars for any given program star. Measurement of relative stellar positions only in azimuth circumvents many of the detrimental effects of Earth's atmosphere on astrometric observations, principally the refractive effects associated with measurement of relative zenith separation. The fully automated operation of the ISI is not a feature unique to the ISI, nor does it offer any strong advantage in terms of better accuracy. It does, however, provide for relatively straightforward real-time internal monitoring and control of the various components of the ISI.

The capability of real-time, on-line data reduction is an important aspect of the ISI. One of the more tedious and time-consuming aspects of most existing photographic astrometry is the postobservational analysis of the plates. The time delay means that if small errors are present in the output data, they are often not discovered in time to obtain substitute observations. Also, the on-line nature of the primary data reduction eliminates many of the sources of error in positional measurement which can arise when data reduction is performed through a sequence of data storage/retrieval processes. Asset (8) is in large measure a consequence of the other items listed above. However, since precision is the ultimate goal of an instrument, it is specifically mentioned here.

Perhaps the best summary statement that can be made as a consequence of this study is that there is no doubt that proper application of modern technology and optics can significantly increase man's ability to study the position and motion of stars. It is somewhat ironical that the type of observational work, which until very recently *was* astronomy, has to a large extent not enjoyed the fruits of modern technology. It is hoped that Project Orion and specifically the concept of the Orion ISI will play a role in redressing this oversight.

1.55-m (61-in.) Reflector
Photograph courtesy of U.S. Naval Observatory.



5. TOWARD A NEW FRONTIER

CONCLUSIONS OF THE ORION STUDY

Since the key detailed technical findings of Project Orion are summarized at the end of each chapter, the conclusions delineated below are of a more general nature. To be sure, many of these conclusions are not new, and many must await the results of further studies before they can be generally accepted. However, in view of the fact that this study constitutes the most in-depth analysis to date of the problem of detecting other planetary systems, there is reason for some confidence in the results.

A search for other planetary systems involves many diverse scientific fields. This involvement ranges from the incidental (i.e., contributions to astronomy unrelated to a search *per se*, but derived from observations made possible with instrumentation developed for a search) to the intentional. Knowledge of the frequency of occurrence and distribution (as a function of, say, spectral type of star) of planetary systems would provide a valuable test of our concepts of the process of star formation. Such knowledge, especially knowledge concerning which *specific* stars have planetary companions, would be very useful (but *not* essential) to any attempt to search for extraterrestrial intelligence. Finally, a search will provide us with perhaps the only means whereby we can test various hypotheses concerning the origin of the solar system.

There exists a rather broad range of observational techniques that might be used to detect other planetary systems. Each of the three techniques considered during Project Orion appears to be feasible in terms of conducting a significant search. There are other potentially useful techniques (e.g., radial velocity observations) that have not been considered here. The nature and scale of technical problems vary markedly among techniques. However, it is important to note that *the attainable precision of ground-based astrometry can be improved by more than an order of magnitude.*

Even if it is decided that the more standard methods (e.g., astrometry and radial velocity) are to be emphasized as search techniques, new instrumentation must be developed and built. Such activity could range from new detectors to new telescopes, with the latter being desirable and likely.

If a search program yields positive detection, and if the proper instrumentation is used to conduct the search, then the information content of the observations could reveal useful data concerning discovered planets, such as planetary mass, temperature, orbit, and, with less certainty, planetary size and rudimentary inferences concerning planetary atmospheres.

A comprehensive search for other planetary systems is timely. This timeliness is of both a technological and philosophical nature. State-of-the-art technology, or soon to be achieved technical advances, is all that is required to undertake a search. There is a need for mankind to open new scientific vistas that offer challenge and excitement not only to those involved in exploring these vistas, but to all mankind. Certainly SETI is one such exploration, and a search for other worlds is one also. Astrophysicist Martin Rees has been quoted as saying that in a SETI endeavor "absence of evidence is not evidence of absence." *However, one of the most exciting aspects of a search for other planetary systems is that once the search is done, we will have definite answers.*

The choice of technique, instrumentation, and magnitude of a search effort is dictated by which of two questions one seeks to answer. If one only wishes to know whether there exist planetary systems around stars close to the Sun, and if he is not concerned with obtaining statistical information on planetary systems as a general phenomenon and all that that statistical knowledge portends for our understanding of the origin of the solar system, a rather modest effort *may* provide the answers. If one wishes to address the more fundamental, quantitative aspects of the problem, ultra-precise instrumentation is required, leading to a larger scale effort. *Even a large-scale effort, involving telescopes in space, is likely to cost no more than a typical planetary spacecraft mission.*

Our final conclusions are self-evident: more studies are needed. There are many facets to a search program – scientific, technological, and programmatic. Studies should be carried out to determine whether the atmospheric limitations to astrometric precision are

those stated here. Further studies of the type carried out by the Greenstein and Black workshops are essential. Finally, because the primary scientific question to be addressed by a search for other planetary systems is the origin of the solar system, it seems appropriate for the NASA to take a leading role in the search.

RECOMMENDATIONS FOR FUTURE ACTIVITY

Some recommendations for future activities related to the search for other planetary systems are:

1. A search program for other planetary systems, with its own budget and funding, should be included in NASA activities.
2. Comparative studies of techniques, both ground-based and space-based, should be funded and undertaken as soon as possible.
3. University and government scientists, both in the United States and in foreign countries, should be made aware of the potential of a search program and be encouraged to participate.
4. If the studies recommended in item (2) indicate that a ground-based program can provide a statistically significant search, steps should be taken to identify a suitable observing site (if no existing facility is suitable).
5. Systems design studies should be funded to identify key technology needs or problems.
6. Any new telescope facilities constructed for a search, whether on the ground or in space, should have ample design input from astronomers who might wish to utilize them for other types of observational studies.
7. Both a *direct* and an *indirect* search technique should be employed in a comprehensive program. Although either could be used alone to carry out a search, the use of both would provide maximum search sensitivity in addition to maximum information return concerning any discovered planets.

The conclusions and recommendations listed above represent a consensus of the Directors and Advisors of Project Orion at the project's conclusion. Although most of the "Orion people" would agree

with many of the conclusions and recommendations it would be misleading to represent the statements in this chapter as a consensus of the entire Design Study Group.

APPENDIX A

VARIABILITY OF OPTICAL PATHS THROUGH THE ATMOSPHERE DUE TO THERMAL TURBULENCE

The turbulence near the surface of Earth is very complex, due in part to wind shear in a boundary layer a few kilometers deep. Thus energy input to the turbulent velocity field involves a variety of distance scales. This kinetic energy cascades to smaller sizes, by a little-understood mechanism of mutual breakup of air parcels not in equilibrium with their surroundings, until viscous dissipation slows the mechanical motions enough for heat conduction to smooth the temperature inhomogeneities. Probably because of the variety of the magnitudes with which energy enters the turbulence, the dissipation rate is inhomogeneous in space. Thus, for example, Merceret (ref. 44) reported that this intermittency in the energy dissipation amounted to two orders of magnitude or more over scales ranging from 100 m to several kilometers in airplane flights at altitudes of 150 and 900 m. The intermittency is manifest in both velocity and temperature fluctuations. At 1-m altitudes, it is evident at centimeter to meter dimensions in the temperature data of Lawrence et al. (ref. 45). But in balloon flights of thermometers to 15 km, Bufton et al. (ref. 46) and Bufton (refs. 47 and 48) found that vertical intermittency did not involve layers more than about 100 m deep.

An important simplification in the problem is that the decay of turbulence involving air parcels small in comparison to the altitude of the parcels closely resembles that in isotropic homogeneous turbulence.

If the steady, smooth flow of air in a wind tunnel is interrupted by a grid, the velocity field is strongly perturbed near the grid but relaxes to smooth flow far behind it. A statistical treatment of the decay of the turbulence was begun by Taylor (ref. 49) and taken up

by many others later. As noted in the review of Corrsin (ref. 50), a major effort has been a search for the shape of the correlation function relating the way in which fluctuations at two nearby points are related (or what is equivalent, the Fourier transform or "spectrum" of the correlation function). In the example of the grid, two points separated by one or more grid spacings tend to be unrelated in all fluctuating quantities such as velocity, density, and temperature. But two points much nearer than a grid spacing will *see* very similar fluctuations. More generally, the complete statistical problem would require knowledge of the mean strength of turbulence and, besides the two-point correlations, all higher N-point correlations as a function of space and time.

The difficulty of the mathematical problem is akin to that in quantum mechanics, as pointed out by Molyneux (ref. 51). Approximation methods are so far essential. Therefore, the approximations for fluctuations of various sizes must be spliced together, using experiment as a guide to the selection of better procedures. Most notably, near the size scale at which energy is being injected into the turbulence, the so-called "outer scale L_0," a useful approximation for the correlation function, is hard to define. Generally, the turbulence departs markedly from both isotropy and homogeneity for point separations larger than about L_0. Von Kármán (ref. 52) showed that, at small wave numbers (large sizes), the spectrum must begin like $(kL_0)^4$. At wave numbers larger than $2\pi/L_0$, the spectrum is proportional to $(kL_0)^{-5/3}$, as found by Kolmogorov (refs. 53 and 54), Onsager (ref. 55), and von Weizsacker (ref. 56). Von Kármán proposed the simplest possible interpolation formula for the spectrum

$$\Phi(B) = \text{const}\ \frac{(k\bar{L}_0)^4}{[1 + (k\bar{L}_0)^2]^{17/6}} \tag{A1}$$

and reported experimental wind-tunnel data that confirmed the usefulness of this formula. Here $\bar{L}_0 = L_0/2\pi$.

In the same spirit, Reiger (ref. 6) noted from the result of Corrsin (ref. 50) that the power spectrum of temperature fluctuations should begin like $(kL_0)^2$ and follow the $-5/3$ law in the inertial

subrange from $2\pi/L_0$ to $2\pi/L_i$. He proposed that the one-dimensional (radial) power spectrum might usefully be taken to be

$$G_1(k) = C_1 \frac{(k\bar{L}_0)^2}{[1 + (k\bar{L}_0)^2]^{11/6}} \tag{A2}$$

Experiments with four thermal probes 2 m above the ground reported by Pasqualetti et al. (ref. 57) generally support the accuracy of the Reiger power spectrum. They find that L_0 = 1 to 3 m in various averages, where 2 m might be expected. Convergence of the averages was so slow that averaging times of 10 to 30 min were necessary. This slow convergence should not be considered surprising in view of the known inhomogeneity of turbulence in the atmosphere. A mathematical model that assumes homogeneity may predict certain parameters of the atmosphere rather well in long-term averages, but the intermittency suggests the caution that events seemingly rare according to the formalism may be not so rare in experiment.

We shall next estimate the "standard error" in optical path lengths through the atmosphere, which will depend on the mean strength of turbulence distributed through the atmosphere. Because the strength of turbulence is itself a random variable of large range, we must guard against any tendency to discuss "two standard error confidence levels" and so on.

When light of known coherence properties enters a turbulent medium, its coherence properties are modified. Like the mathematical problem of characterizing the turbulence, the statistical problem of describing the wave field has no known exact solution. Approximation methods are again resorted to, with much criticism of the suggestions about the domain of validity of the approximations between various workers. In a review of approaches to a solution of the relevant nonlinear stochastic differential equation, Fried (ref. 58) has remarked, "If it were not so well known that light does somehow manage to propagate in the atmosphere, we might very well have heard questions raised concerning the existence of a solution!"

The intended use of the present study is primarily a feasibility study for interferometric measures through the atmosphere. In the Michelson stellar interferometer, for example, light from two finite areas is combined so as to study the correlation of the electric fields

arriving at the two apertures. The apertures are finite so each aperture averages over the small details of the electric field distribution that might be present. If we could follow the path of the incident light beams through the atmosphere to the final defining apertures, we could, in principle, calculate what details of the index of refraction transverse to the beams are thereby averaged out. This problem has been treated by Cook (ref. 59) by a method that gives a correct wave-optical treatment for the centroid of a finite-width beam; the treatment is valid over long propagation paths. He uses Ehrenfest's theorem to transfer a result established in quantum theory to the macroscopic wave-optical domain: the average position of a particle is calculable from Newton's second law of motion when the force is taken to be the average value of the negative gradient of the potential. The analogous centroid of the beam is the first moment of the intensity. He shows that this beam centroid is governed by the differential equation for a paraxial ray in a weakly inhomogeneous medium, provided that the true refractive index over the beam is replaced by a smoothed version, where the smoothing function is the intensity profile of the beam.

At a site of low natural turbulence and at times of especially low turbulence in the atmosphere, which would be selected for experimental purposes, the motion of the beam centroid can be estimated from the formalism of Cook. The same formalism gives an estimate of the variance of the beam angle of arrival. Without going into details about assumptions concerning the turbulence distribution until later, the turbulence that gives angles of arrival mostly from a 1.0-arcsec range of directions also causes the beam centroid to wander by ±0.5 cm. Since the effective center of the beam then moves an order of magnitude less than the size of each aperture of interest in a Michelson stellar interferometer, we shall be well justified in using a ray-optics formalism. Furthermore, because the inner scale in the first 10 km of the atmosphere is small relative to the interferometer apertures, we need not concern ourselves with modifications of the power spectrum due to viscous effects.

The above assumptions are just those necessary for the validity of the calculations of Hufnagel and Stanley (ref. 60, here called HS), whose notation we shall follow. Introduce the vector notation:

$$\vec{r} \equiv (x,y,z) , \qquad \vec{p} \equiv (x,y)$$

The wave equation for the propagation of a scalar V through a lossless nonhomogeneous medium is

$$\Delta V - (n/c)^2 \partial^2 V/\partial t^2 = 0 \tag{A3}$$

where Δ is the three-dimensional Laplacian operator, c is the speed of light in a vacuum, and $n = n(\vec{r},t)$ is the local index of refraction. The normalized fluctuating part of n is defined by

$$N(\vec{r},t) \equiv \frac{n(\vec{r},t) - n(\vec{r})}{n(\vec{r})} \tag{A4}$$

where $\bar{n}(\vec{r},t)$ is the local time average of n. Without noticeable loss of accuracy, we take $\bar{n}(\vec{r}) = 1$. If N is everywhere zero, a solution of equation (A3) corresponding to a plane wave propagating in the positive z direction is

$$V(r,t) = A \exp[i(kz-wt)] \tag{A5}$$

where $k = 2\pi/\lambda$, λ is the wavelength of light, $w = kc$, and A would be a constant. But in the presence of nonzero $N(\vec{r},t)$, A is not constant but responds to the index-of-refraction variations according to

$$\Delta A + 2ik\, \partial A/\partial z + 2k^2 NA = 0 \tag{A6}$$

We shall be interested, however, in the average coherence function

$$M(\vec{r}) = \langle V(\vec{p}' + \vec{p},z,t) V^*(\vec{p}'z,t) \rangle$$

$$= \left\langle \exp\left\{ ik \int_0^h [N(\vec{p}_1,z) - N(\vec{p}_2,z)]\, dz \right\} \right\rangle \tag{A7}$$

according to *HS*. Here h is the altitude of the tropopause for a vertical integration path. Indeed, our central interest is in the difference in two optical paths separated by B:

$$L_2(\vec{B}) = \int_0^h [N(\vec{p},z) - N(\vec{p} + \vec{B},z)]\, dz \tag{A8}$$

The mean value for L_2 vanishes as an ensemble average, which we shall take to be a long-term average. However, the variance in L_2 is calculated from

$$\langle L_2{}^2 \rangle = \int_0^h dz_1 \int_0^h dz_2 \langle [N(\vec{p}_1, z_1) - N(\vec{p}_2, z_1)]\,[N(\vec{p}_1, z_2) - N(\vec{p}_2, z_2)] \rangle \tag{A9}$$

At this point, we introduce the covariance function for the random field N:

$$\langle N(\vec{r}_1) N(\vec{r}_2) \rangle = a^2 C(|\vec{r}_1 - \vec{r}_2|) \tag{A10}$$

and the distances $z = z_2 - z_1$, $B = |\vec{p}_2 - \vec{p}_1|$, whence

$$\langle L_2{}^2 \rangle = 2 \int_0^h dz_1 a^2(z_1) \int_{-z}^{h-z} dz\,[C(z) - C(u)] \tag{A11}$$

where $u^2 = z^2 + B^2$. We have indicated that the covariance parameter a will generally be a slow function of altitude. On the other hand, the covariance function C vanishes rapidly when the distance between two points is larger than the outer scale L_0, which is always much smaller than the tropopause height h. We are then led to introduce normalized functions F that are line integrals over the covariance function at a "miss distance" x:

$$2bF(x) = \int_{-\infty}^{\infty} C[(z^2 + x^2)^{1/2}]\,dz \tag{A12}$$

where b is chosen so that $F(0) = 1$. With no substantial error, we may approximate equation (A11) as follows:

$$\langle L_2{}^2 \rangle = 4 \int_0^h dz\, a^2(z)[1 - F(B)]\, b(z) \tag{A13}$$

We know from the flights of Bufton that the z (altitude) dependence of $a^2(z)$ is usually approximately exponential with a scale height of about 4 km. The length $2\pi b$ is roughly the outer scale L_0, which is known from many experiments, including that of Pasqualetti et al. (ref. 57), to vanish linearly at ground level. There seems to be no other experimental evidence on the altitude dependence of the outer scale except that it is not likely that the outer scale could exceed the layer thicknesses found by Bufton. A suitable model for the outer scale for the present purpose is a linear increase $L_0(z)$ up to a saturation value that holds for the remainder of the atmosphere above altitude h_1. The function $F(x)$ can be evaluated and stored. If the integral equation (A13) is done in units of L_0', the saturation value of L_0, then $F(B)$ can be evaluated by the scaling relaxation $x = BL_0'/L_0(z)$. On the other hand, we shall argue below that L_0' is on the order of 100 m, less than 1 percent of h, so it is not likely that any great error can be made by taking the outer scale constant all the way to the ground. In that case, $b[1 - F(B)]$ are not functions of altitude and the remaining integral over altitude gives

$$\langle L_2{}^2 \rangle = 4b[1 - F(B)]Ha^2(0)[1 - \exp(-h/H)] \qquad \text{(A14)}$$

where H is the exponential scale height for $a^2(z)$.

The fact that the two-aperture optical path difference L_2 has a variance that is roughly $2L_0' \cdot 2H \cdot a^2(0)$ at large B, where F must vanish, suggests a simple interpretation. The effective path length for two paths is $2H$ at a turbulence strength of $a^2(0)$. If we had to estimate the variance in L_2 from $2H/2L_0'$ regions that are independent of each other, each of size $2L_0'$, we would calculate

$$(2L_0' a)^2 \cdot (2H/2L_0')$$

which is just the present result.

Suppose we add the idea that not all of the independent regions are at the same stage of turbulent decay; then we might even expect that the turbulence strength a^2 in different regions will appear to be variable. Lin (ref. 61) pointed out evidence from wind-tunnel data which called for a dissipation rate that was an explicit function of time after the energy injection. The unpleasant aspect of his theory for the two- and three-point correlation functions is that, at the

beginning of the decay process, the functions are self-preserving in shape only at small r/L_0, whereas late in the decay process the whole curve is fixed in shape. This means that the thermal spectrum is apt to deviate from equation (A2) in a manner that depends explicitly on the current value of a^2 observed in the same region. A tendency of this sort is clearly visible in the airplane flight data of Merceret (ref. 44). In Merceret's figure 6 (ref. 44), for example, there is a clear tendency for high-turbulence power spectra to follow the $k^{-5/3}$ law. But the low-turbulence spectra follow it only at high frequencies. At low frequencies, the deviations from the $k^{-5/3}$ law are in the sense of the interpolation formula (A2). In terms of the concepts suggested by Lin (ref. 61) and von Kármán (ref. 52), these data suggest that the higher turbulence regions be thought of as longer developed regions so that both the accuracy of the $-5/3$ law has improved and the total energy transferred into the high wave numbers by action of Reynolds stresses has had time to increase.

For present purposes, we do not need a highly accurate shape for the two-point correlation function $C(r)$ in equation (A12) because of the line integral with respect to altitude. We shall use equation (A2) for the power spectrum of the temperature fluctuations and try to choose L_0 from experiment so as to average the corresponding correlation functions $C(r/L_0)$ along the line of sight over all the states of turbulence development.

The three-dimensional power spectrum corresponding to $G_1(k)$ is (ref. 6):

$$G_3(k) = \frac{G_1(k)}{4\pi k^2} \tag{A15}$$

from which (ref. 62)

$$C(r) = \iiint_{-\infty}^{\infty} G_3(k)\exp(i\vec{k} \cdot \vec{r})d^3k \tag{A16}$$

Using $C(-r) = C(r)$ to drop the imaginary part of the exponential and introducing spherical coordinates in k-space so that

$$d^3k = k^2 \sin\theta \, d\theta \, d\phi \, dk'$$

and

$$\vec{k} \cdot \vec{r} = kr \cos \theta$$

we can perform the ϕ integral immediately. The polar angle integration requires use of the identity:

$$\exp(iz \cos \theta) = \sum_{m=0}^{\infty} (2m + 1) i^m j_m(z) P_m(\cos \theta) \tag{A17}$$

where the spherical Bessel functions $j_m(z)$ are regular at $z = 0$ and $P_m(\cos \theta)$ are Legendre polynomials. Fortunately, the integral over θ vanishes for all but $m = 0$:

$$\int_0^{\pi} P_m(\cos \theta) \sin \theta \, d\theta = \begin{cases} 2 \text{ for } m = 0 \\ 0 \text{ for } m > 0 \end{cases} \tag{A18}$$

Thus we find

$$C(r) = \int_0^{\infty} \frac{\sin kr}{kr} \, G_1(k) dk \tag{A19}$$

Because $k^{-1} G_1(k)$ vanishes at $k = 0$ and ∞, an integration by parts and a substitution $u = k\bar{L}_0$ gives

$$C(\rho) = \frac{3}{5} \frac{C_1}{\bar{L}_0} \int_0^{\infty} \frac{\cos \rho u \, du}{(1 + u^2)^{5/6}} \tag{A20}$$

where $\rho = r/\bar{L}_0 = 2\pi r/L_0$. The constant C_1 is fixed by our desire to have $C(0) = 1$, so that

$$C_1 = \frac{5}{3} \frac{\Gamma(5/6)\bar{L}_0}{\Gamma(3/2)\Gamma(1/3)} = 0.79241 \, \bar{L}_0 \tag{A21}$$

From tables of cosine transforms (ref. 63)

$$C(\rho) = \frac{3}{5}\frac{C_1}{\bar{L}_0}\frac{\pi^{1/2}}{\Gamma(5/6)}\left(\frac{\rho}{2}\right)^{1/3} k_{1/3}(\rho) \tag{A22}$$

where

$$K_\nu(z) = \frac{\pi/2}{\sin \pi\nu}\left[I_{-\nu}(z) - I_\nu(z)\right]$$

is defined in terms of the modified Bessel function:

$$I_\nu(z) = \frac{(z/2)^\nu}{\Gamma(\nu+1)}\left[1 + \frac{(z/2)^2}{(1)(\nu+1)} + \frac{(z/2)^4}{(1\cdot 2)(\nu+1)(\nu+2)} + \ldots\right] \tag{A23}$$

A useful series for small ρ is

$$C(\rho) \cong (1 - R\rho^{2/3}) + \frac{(\rho/z)^2}{1!}\left(\frac{1}{2/3} - \frac{R\rho^{2/3}}{4/3}\right) + \frac{(\rho/2)^4}{2!}\left(\frac{1}{\frac{2}{3}\cdot\frac{5}{3}} - \frac{R\rho^{2/3}}{\frac{4}{3}\cdot\frac{7}{3}}\right) + \ldots \tag{A24}$$

where

$$R = \frac{3}{2^{5/3}}\frac{\Gamma(5/3)}{\Gamma(4/3)} \cong 0.955273$$

At large ρ, an asymptotic expansion for $C(\rho)$ is

$$C(\rho) \sim \frac{2^{1/6}\pi^{1/2}}{3\Gamma(4/3)}\frac{e^{-\rho}}{\rho^{1/6}}\left[1 + \sum_{m=1}^{M}\frac{(-1)^m(1^2 - 4/9)\ldots[(2m-1)^2 - 4/9]}{m!(8\rho)^m}\right] \tag{A25}$$

where, in the sum to M terms, it is convenient to take terms by pairs and stop when the contribution is either smaller than some desired precision or large enough to make $C(\rho)$ increase. In a computer carrying 16 binary bits, it was found that a crossover point between large and small ρ at $\rho = 2.883400$ gave results with minimum discontinuity at the crossover. The function $C(\rho)$ is 0.34 at the crossover point and decreases exponentially with ρ.

When $C(\rho)$ is given by equation (A22), the line integral giving b in equation (A12) can be obtained analytically as a K transform (ref. 63):

$$b = \bar{L}_0 \int_0^\infty C(\rho)d\rho = \bar{L}_0 \frac{\Gamma(5/6)\Gamma(1/2)}{\Gamma(1/3)} \cong 0.74683\, \bar{L}_0 \qquad \text{(A26)}$$

The function $F(x)$ was obtained numerically, using equation (A26) to verify the accuracy of the procedure. A file of values ranging over six orders of magnitude was stored for regularly spaced arguments, spaced by $\bar{L}_0/50$. Then the function $F(x)$ at arbitrary argument could be obtained rapidly by interpolation.

When the baseline B is very much smaller than the outer scale L_0, we can use equation (A24) to find the limiting behavior of $F(x)$ as $x = 2\pi B/L_0$ goes to zero. If $u^2 = z^2 + x^2$,

$$b[1 - F(x)] = \int_0^\infty dz[C(z) - C(u)] \cong R \int_0^\infty dz[(z^2 + x^2)^{2/3} - z^{2/3}]$$

$$= Rx^{5/3} \int_0^\infty dy[(y^2 + 1)^{1/3} - y^{2/3}] \cong 1.392\, x^{5/3} \qquad \text{(A27)}$$

Even at $B = L_0/50$, the slope of $\log[1 - F(x)]$ versus $\log x$ was not 5/3, but 1.40. A detailed study at small x was not carried out, but the interpolation formula for $F(x)$ was taken to be of the form indicated by equation (A27), with the coefficient 1.392 replaced by 1.513 to match the function at $L_0/50$.

The result of the procedure gives the rms values for L_2 calculated from equation (A13) shown in figure 7. We note that the average L_2 saturates at a baseline equal to the outer scale. At small baseline, the variance $L_2{}^2$ exhibits a power-law dependence on baseline that gradually steepens to the 5/3 slope arranged by the interpolation procedures discussed above. The inclusion of a region of decreased outer scale next to the ground smooths out the L_2 curve so that no sharp break in slope is evident at $B = L_0/50$. The L_2 values assume that the turbulence strength has an exponential scale of 3.75 km until an altitude of 16 km, where turbulence is taken to vanish. At the ground, we take $a^2(0) = 4\times10^{-16}$ $m^{-2/3}$ from the data of Bufton et al. (ref. 46), which correspond to quiet conditions at dawn. The wavelength $\lambda = 0.5$ μm is chosen for example only.

The experimental results of Bouricius and Clifford (ref. 64) confirm departures from the 5/3 law of equation (A27). They studied the variance in phase measures in a horizontal light beam experiment about 2 m above ground. They found a slope near 1.5 and looked carefully for possible reasons for a discrepancy with the expected 5/3 slope. It seems not to have been considered that the effects of proximity to the saturation at $B = L_0$ are evident to point separations of 10^{-2} L_0. The result is confirmed in the calculations of Cook (ref. 59), who studied the variation in angle of arrival of a finite beam. His result is important for a conclusion from the vertical placement of the curves in figure 7.

According to Cook, the variance of angle of arrival for a collimated beam of size T (which he takes to have Gaussian intensity distribution) is of the form

$$\langle\alpha^2\rangle = C_2 a^2 z L_0{}^{-1/3} f(z) \tag{A28}$$

where the dimensionless variable $x^2 = (T^2 + 4L_i{}^2)/L_0{}^2$ can be taken to be $x = T/L_0$ in the present context since the inner scale L_i is less than centimeter size near the ground. The constant C_2 is 1.303. The combination za^2 is to be interpreted as the integral over turbulence strength as in equation (A13). The function $f(x)$ is very closely related to $[1 - F(B)]/B^2$, which, according to equation (A27), is proportional to $B^{-1/3}$ at small baseline B. Clearly, L_2/B has the significance of a wave-front tilt like α. Cook gives

$$f(x) = 3.51x^{-1/3} - 3.60 + 1.93x^{5/3} - 1.54x^2 + \ldots \qquad \text{(A29)}$$

from which the first term and the assumptions of figure 7 give $\alpha = 0.536$ arcsec at $T = 1$ m. But Fried (ref. 65) has pointed out that when imaging through turbulence described by the 5/3 law of equation (A27), an overwhelming part of the deformation of the wave front is simple tilting, especially for long-term exposures. Thus we may say that the assumptions underlying figure 7 imply that a star image at the zenith would be about 2α, or about 2.2 arcsec in angular diameter. The dependence of this result on telescope size T is only an inverse 1/6 power, which is too slow to notice as an explicit aperture dependence on the seeing disk size. The dependence on outer scale drops out completely, unless the second and later terms in equation (A29) are substantial. For example, $[x^{1/3}f(x)/3.51]^{1/2} = 0.88$ for $T = 0.01\ L_0$, which lowers 2α to 1.0 arcsec.

The use of direct imaging to study the product za^2 under known micrometeorological conditions has been used, for example, by Wesely and Derzko (ref. 66). They find good agreement between limiting resolution obtained visually through 8- to 15-cm telescopes and measured za^2 products inserted into a formula comparable to equation (A28). The theory is due to Fried (ref. 67), who introduced the length D through the relation

$$(2\pi L_2/\lambda)^2 = 6.88(B/D)^{5/3} \qquad \text{(A30)}$$

and showed that the limiting resolution of a system not able to respond rapidly to the wave-front tilts could be taken to be equal to the diffraction-limited resolution of a telescope having diameter D. This was confirmed by Wesely and Derzko, who measured the Rayleigh criterion angle

$$\theta = \frac{3.34}{\lambda^{1/5}} (za^2)^{3/5} \qquad \text{(A31)}$$

Actually, various coefficients close to 3.34 could be adopted because of the visual determination of the limiting resolution. It is possible, however, that finite outer scale effects were detected because, as we saw above, θ tends to decrease if T/L_0 is significant. A major part of

the data was taken with $T = 15$ cm and $L_0 \cong 1$ m, for which θ might be decreased by a factor of 0.66. The optically inferred za^2 product was either 0.77 or 0.82 times the value calculated from temperature and water vapor fluctuations, depending on one's assumption about the correlation between temperature and water vapor. It is also interesting to note from equation (A31) that wavelength dependence of the seeing spot size is almost absent, in agreement with the absence of color effects in the speckle pattern of a large telescope.

In a paper of crucial interest to the interpretation of figure 7, Breckenridge (ref. 68) reported the fringe motion in a wave-front folding interferometer at conditions of known seeing. He reported the total range of fringe motion which was exceeded no more than 10 percent of the time when the seeing disk was 3 arcsec across. Assuming Gaussian statistics are accurate enough for the present purpose, we converted this total range to a standard error by dividing by 3 and plotted his results in figure 7. Breckenridge made the point that the placement of his points was consistent with equation (A30), that is, $L_2{}^2$ was proportional to the 5/3 power of B. The points are rather better fit by a curve for which 5/3 is replaced by 1.5, as in the results of Bouricius and Clifford (ref. 64) discussed above. At $B = D$, we find $\log(L_2/\lambda) = -0.379$ for whatever slope is appropriate in equation (A30). From figure 7 we read that $D = 11.5$ cm, in perfect agreement with the maximum distance in the pupil for which Breckenridge could discern fringes in a long-term photograph of the fringes. But for this value of D, the Rayleigh criterion angle $1.22\lambda/D$ is 1.32 arcsec, in sharp disagreement with the 3-arcsec seeing disk.

The issue posed by this disagreement is that, if seeing were at the 3-arcsec level as he asserted, then the outer scale in most of the troposphere is between 50 and 100 m; that is, the Breckenridge points should be shifted down by a factor of 3 to be on the same basis as the theoretical curves. If seeing somehow were 1 arcsec after all, his data require an outer scale nearer to 10 m.

It is easy to argue that the discrepancy in "resolution angle" is to be expected. According to relation (A2) and for either choice of the outer scale, optical effects grow stronger for heat waves of increasing dimension for baselines in the 1-m range. Larger waves are carried past the telescope in longer times. According to the results of KenKnight (ref. 69), it is unusual that the fringes at more than 0.5-m separation could even be followed visually. Of course, it is to be

expected that the fringes pause briefly at the extremes of their motion. Thus the visual interferometer data heavily weight the extreme fluctuations due to low wave numbers.

On the other hand, the outer extremes of the seeing disk arise from diffraction from the highest wave-number heat waves in front of the telescope. But from figure 7 we saw that L_2/B is proportional to $B^{-1/6}$, which diverges at small B because we have ignored the effects of the inner scale.

More fundamentally, the presence of this divergence is due to the fact that, when imaging with a continuous ("filled") aperture, ray separations of arbitrarily small amount are of interest. In the passage from discrete-ray separations to continuous-ray separations, we are led to consider not only the correlation of index-of-refraction fluctuations, as in equation (A10), but the correlation of gradients in the index of refraction (ref. 70):

$$Q_{ij} = \left\langle \frac{\partial N}{\partial \vec{r}_1} \frac{\partial N}{\partial \vec{r}_2} \right\rangle = -a^2 \frac{\partial^2 C}{\partial \xi_i \partial \xi_j} , \qquad \xi_i = (\vec{r}_1 - \vec{r}_2)_i \tag{A32}$$

Taking advantage of the average isotropy of the medium,

$$\left\langle \frac{\partial N}{\partial x_1} \frac{\partial N}{\partial x_2} \right\rangle = -a^2 \frac{1}{z} \frac{\partial C}{\partial z} \tag{A33}$$

we see that the correlation of wedge effects depends on the distance along the beam z through the transverse correlation function $Q(r)$, where

$$Q(r) = -\frac{1}{r} \frac{dC}{dr} \tag{A34}$$

If we try to obtain Q from the spectral representation (A2) as in (A15) to (A20), we fail because the integral in equation (A20) does not converge at large wave numbers when, in obtaining $C'(r)$, we replace $\cos(ru)$ by $-u \sin(ru)$ in the integrand. Various mathematical devices have been used to effect convergence. For example, Cook (ref. 59) used the modification of equation (A2) associated with the

name of von Kármán, which was a simple multiplication of equation (A2) by $\exp[-(u\bar{L}_i)^2]$, $\bar{L}_i = L_i/2\pi$. This ruse has no physical justification. Corrsin (ref. 50) pointed out that the thermal fluctuation spectrum must steepen from the $u^{-5/3}$ relation in the inertial subrange to u^{-7} in the viscous region. In the spirit of Reiger (ref. 6), an interpolation formula of simplest form is

$$G_1(k) = C_3 \frac{(k\bar{L}_0)^2}{[1 + (k\bar{L}_0)^2]^{11/6} [1 + (k\bar{L}_i)^2]^{5/3}} \tag{A35}$$

The point for the present is that direct imaging tends to weight heavily the smaller heat waves in front of the telescope. Because very many of them project their phase perturbations on the entrance pupil, we realize that we deal with the statistics of very large numbers when we measure the extent of the seeing disk. Therefore, the size of the seeing disk is a relatively robust measure of the turbulence along the line of sight. In contrast, the centroid of a star image, like the phase difference in an interferometer at a separation B approaching the telescope size T, is mainly perturbed by a few large heat waves in the line of sight. The heat-wave size may far exceed T, so it will convect past for a long time.

We therefore accept the conclusion that figure 7 reveals an outer scale in most of the troposphere of 50 to 100 m. The important corollary conclusion is that *the rms difference in two vertical optical paths will not exceed about 2 μm during quiet atmospheric conditions, no matter how far the vertical paths are separated.* At zenith angle ξ, the path difference will be less than $2 \sec^{1/2} \xi$ μm for modest zenith angles. Note, of course, that the prediction ignores density fluctuations associated with mechanical motions such as mountain waves, weather fronts, and convection cells larger than 100 m.

APPENDIX B

ANALYSIS OF IMAGING STELLAR INTERFEROMETER DATA TO DETERMINE RELATIVE STAR POSITIONS

In concept, the imaging stellar interferometer (ISI) blends the advantages of focal plane astrometry with the strengths of long-baseline interferometry. In this section we will develop the rigorous transformations and show their application in a now standard statistical reduction technique. We will present here only the rudiments of the proposed algorithm.

As a region transits one of the two instrumental vertical circles, the focal plane detectors of the respective ISI determine a series of relative azimuths for the target star and approximately 20 selected reference stars. To lessen the effects of atmospheric turbulence and scaling, each detector integrates its positional and intensity measurements over exactly the same increment of time. Thus, a list of relative positions is produced for each object over a series of points in time. The rotation of Earth, its orbital motion, and the space motions of the Sun and stars will make each set of observations unique. However, the equations of motion are known, and rigorous transformations are available to bring all observations into a common reference frame.

Two sets of constraints must be satisfied simultaneously by the statistical reduction of the measured reference star positions. The first models the projected position of these stars to the best existing set of predicted coordinates. The second constrains the apparent motion of these objects to either a linear or a Keplerian trajectory. With rare exception, known binaries will be avoided as reference stars but, as pointed out elsewhere, at high astrometric precision new discoveries will be common. The reduction process is iterative, with each set of measurements being reduced to the positions predicted by the models and parameters that best satisfied all other observations on the previous pass. The algorithm is essentially that of the central overlap technique (ref. 3). However, the geometry is different.

Figure 62 illustrates the celestial sphere of the observer. Both Earth and the observer are assumed to be of insignificant size and to be placed at the exact center of the sphere. As Earth rotates around line SCP–NCP, a star with a declination δ will approach the first vertical circle (FV) at an angle θ:

$$\theta = \frac{\pi}{2} - \arcsin\left(\frac{\cos\phi \sin FA}{\cos\delta}\right) \qquad \text{(B1)}$$

where θ is in radians, ϕ is the latitude of the site, and FA is the azimuth (east of south) of the first vertical circle. A similar equation

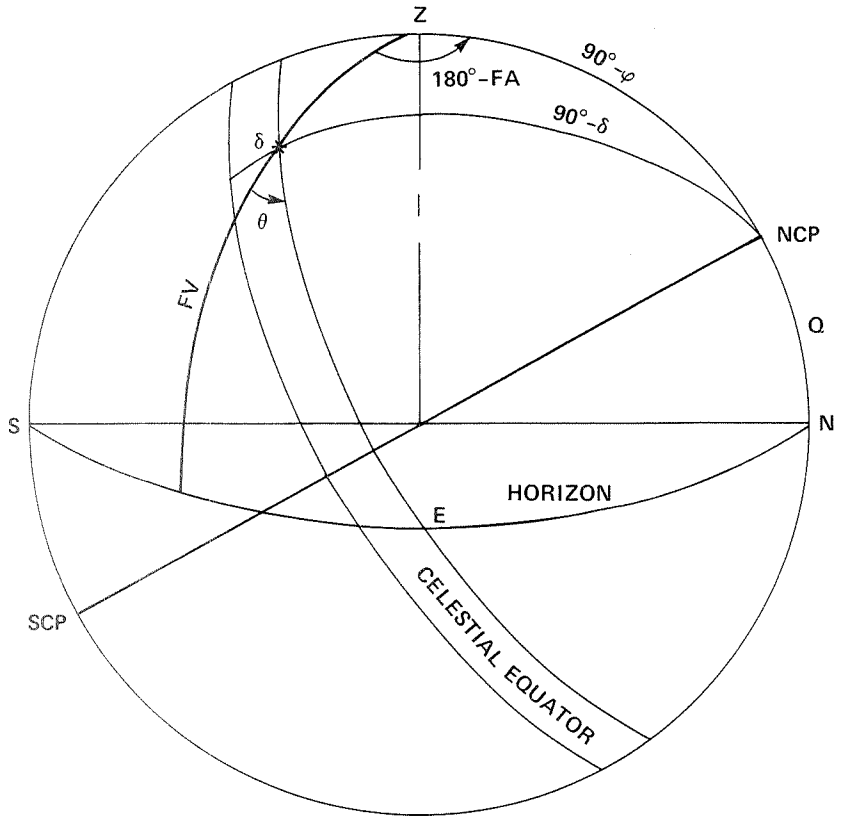


Figure 62.– *Schematic representation of the celestial sphere showing the principal angles used in analyzing the data produced by the Orion ISI.*

exists for the second vertical circle. The dependence of θ on δ causes the field of an ISI to appear to rotate slightly during a transit. But this effect is quite small for low latitudes and a reasonable range of zenith distances and, if we use short periods of simultaneous integration, could be ignored even near the pole. Table 11 lists θ over a 70° arc. Near maximum sky coverage for a site within the United States is achieved by the choice of ϕ and FA utilized there.

The central overlap technique determines the initial scale and orientation of a field through stellar positions and proper motions obtained from catalogs and wide-angle astrographic plates. These positions, corrected for epoch and equinox, may be placed into the altitude (a)-azimuth (A) system of the ISI via the rigorous transformations:

$$a = \arcsin[\cos\phi \cos\delta \cos(T - \alpha) + \sin\phi \sin\delta] \qquad \text{(B2a)}$$

$$A = \arctan\left[\frac{\sin(T - \alpha)}{\sin\phi \cos(T - \alpha) - \cos\phi \tan\delta}\right] \qquad \text{(B2b)}$$

where a and A are to be converted from radians, α and δ are the right ascension and declination of the object, and T is sidereal time. If the denominator equation (B2b) is positive, we add 180° to A; if it is negative and the numerator is positive, we add 360°. Equations (B2) are valid for either ISI.

Each ISI images a region onto its focal plane where the relative positions are measured. This plane is similar to that which contains the photographic plate or multichannel astrometric photometer (MAP) of a more standard astrometric instrument; the coordinates of the positions of objects on the celestial sphere imaged onto the focal plane are similar to the standard coordinates used in that connection.

TABLE 11.– FROM EQUATION (B1) ASSUMING $\phi = 20°, FA = 30°$

δ, deg	θ	δ, deg	θ
−15	60°53′	30	57°08′
0	61°58′	45	48°21′
15	60°53′	55	35°00′

Where (A_0, a_0) are the azimuth and altitude of the point at which the optical axis of the respective ISI intersects the celestial sphere, we have the rigorous transformations:

$$\beta = -\frac{\cos a \sin(A - A_0)}{\sin a \sin a_0 + \cos a \cos a_0 \cos(A - A_0)} \tag{B3a}$$

$$\gamma = \frac{\sin a \cos a_0 - \cos a \sin a_0 \cos(A - A_0)}{\sin a \sin a_0 + \cos a \cos a_0 \cos(A - A_0)} \tag{B3b}$$

where the β axis is parallel to the horizon, increasing with decreasing north azimuth, and the γ axis contains the projection of either FV or SV. The tangential coordinates (β,γ) are given here in radians. Because the Orion system does not track the field under observation, the equatorial coordinates of the tangent point change continuously. Thus the tangential coordinates of each star must be recomputed for each set of observations, approximately 300 times during one observing session. However, the number of computations necessary for 20 or so stars in these and all other involved transformations will hardly challenge the speed of a modern minicomputer.

The predicted instantaneous tangential coordinates now form a lattice for the adjustment of those obtained with the respective ISI during a given observational increment. The effects to be modeled and removed via this adjustment are more numerous than is immediately obvious. By this reduction we will place a set of measured coordinates into a reference frame defined by a subset of the stars which comprises a standard catalog. The zero point and orientation of our reduced positions will be essentially that of the catalog system at the equator and equinox to which we have precessed these positions. The epoch will be that of the observation. The effects of nutation will be removed by a field translation as will those of general refraction. The first-order effects of stellar aberration will be removed by scaling the field to match the predicted positions as will those of differential refraction. Errors of mechanical alignment on the instrumental vertical circle and pivot errors as well as errors in assumed longitude and variations in geographic latitude will be

removed by an affine transformation. Nongnomonic projection characteristics of the system will require higher order terms in each coordinate as will variations in differential refraction across the field. Probably the major source of noise in the measured positions will be the rapidly changing refractive characteristics of Earth's atmosphere (ref. 28). Unmodeled, these variations will cause apparent displacements, scale changes, and distortions which combine to introduce positional errors thousands of times larger than the theoretical precision of a stellar interferometer.

If we express the effects of each of the above phenomena on the stars within the region defined by the field of an ISI in a set of Taylor series expansions, truncate those expansions, and then add them together, we may write:

$$\beta = bx + c\gamma + d + ex^2 + fx\gamma + g\gamma^2 \qquad \text{(B4)}$$

where x is the measurement provided by the detectors of the relevant ISI with $x = 0$ as the detectors cross the projected vertical circle near the optical axis of the instrument. The quantities β and γ are predicted by equation (B3), and the parameters b through g are to be determined via least squares.

There is some uncertainty about the point of truncation in the relevant Taylor series. It is quite possible that the misbehavior of the atmosphere will require a general third-order expression and hence 10 terms in equation (B4). Special terms may be necessary to model the nongnomonic imaging caused by those optical surfaces that the design has not been able to place in the optical pupil of the system. Magnitude or color-related displacements, magnifications, or distortions may be present. To ensure reasonable stability in the statistical reductions, each new term added to equation (B4) should be accompanied by the measurement of approximately three new reference stars. As with any precision instrument, a rash of such unexpected terms could pose serious questions about the system as a whole.

As a region crosses the field of an ISI, the imaged stars appear to move along a set of gentle arcs which are the projections of small circles onto a tangent plane. Thus, the x positions recorded for a star during its transit move rapidly across the field and that motion suffers a spurious acceleration. However, we may write equation (B4) in the form:

$$\beta - (b_0 x + c_0 \gamma + d_0) = b'x + c'\gamma + d' + ex^2 + fx\gamma + g\gamma^2 \qquad \text{(B5)}$$

where b_0, c_0, and d_0 are initial estimates of the rotation and translation parameters and b', c', and d' are corrections to these estimates to be calculated by statistical adjustment. The constants e, f, and g will be of the order of b', c', and d' and can be estimated directly with no loss of generality or precision. The right-hand side of equation (B5), unlike that of (B4), shows essentially only the trends we seek to explain. Furthermore, these trends can now be directly compared with the parameters of time and symmetry with the optical axis.

The first of these, time, will allow the investigator to model the passage and variation of seeing phenomena from one second to the next. Experience will probably suggest certain constraints that can be applied to these wavelike phenomena. The motion of the region across the field should allow the observer to detect symmetrical imaging errors of the system by employing the concept of overlap condition (ref. 71). This concept will also allow stars spread over an area somewhat larger than the field of the ISI to be utilized as reference points.

The advantages of this type of relative adjustment have long been recognized in astrometry. Conditional equations similar to equation (B5) with less than a dozen parameters adequately model the various effects that all phenomena discussed previously can exert on the target star's position in a reference frame defined by those stars that lie nearly in the same direction. Similar reductions into an absolute system of coordinates would require scores of parameters with their associated parameter variance. To gain some of the precision of focal plane astrometry, instruments that measure the spherical coordinates of one star at a time sometimes follow a procedure of alternately observing the target star and a few reference stars. However, as pointed out by Schlesinger (ref. 72) and Hudson (ref. 73), and as evidenced by the work of KenKnight (ref. 28), *simultaneous observations are absolutely necessary.* Furthermore, at the precisions sought, one must remove at least the effects of first- and second-order differential refraction. *Such a reduction includes the determination of at least 6 coefficients and the simultaneous observation of perhaps 18 or 20 reference stars.*

There is one additional form of astrometric data provided by the detectors. Each unit determines the position of a star in four different bandpasses. The refractive indices of these four colors have a range of approximately 1/50 their mean magnitude. Thus a trend in position with wavelength should be measurable. This trend allows an estimate of the unrefracted position of the star. However, the trend, including both the accidental and the systematic errors of the observation, must be multiplied by 50 to estimate the correction necessary to remove the effects of instantaneous refraction. Fortunately, this information is already available at relatively high weight in the simultaneously observed positions of numerous reference stars.

Once a satisfactory reduction of the measured coordinates has been effected, we turn to the second major phase of an iteration, an analysis of the parameters of motion of each star. As currently applied at this point in the algorithm, the central overlap technique transforms the observations of each star into a different planar coordinate frame. Tangent to the celestial sphere near the center of the star's great circle trajectory with projected displacements recorded in standard coordinates, this reference frame is unusually free of spurious accelerations and is adequate for several decades of analysis without correction. We will incorporate one additional rotation in this plane into our transformation so that each star's motion is modeled first in the coordinates obtained with one ISI and then in terms of those obtained with the other ISI.

The four sets of rigorous transformations used to place the data in the desired coordinate frames begin by converting the observed tangential azimuths to observed altitudes and azimuths:

$$A = A_0 + \arctan[-\beta/(\cos a_0 - \gamma \sin a_0)] \tag{B6a}$$

$$a = \arcsin[(\sin a_0 + \gamma \cos a_0)/(\beta^2 + \gamma^2 + 1)^{1/2}] \tag{B6b}$$

and rigorous formulas:

$$\alpha = \arctan\left[\frac{\sin T(\cos\phi \sin a - \sin\phi \cos a \cos A) + \cos T \cos a \sin A}{\cos T(\cos\phi \sin a - \sin\phi \cos a \cos A) - \sin T \cos a \sin A}\right] \tag{B7a}$$

and

$$\delta = \arcsin(\sin\phi \sin a + \cos\phi \cos a \cos A) \tag{B7b}$$

where α and δ are in radians. If the denominator of equation (B7a) is negative, we add 180° to α. If the denominator is positive and the numerator is negative, we add 360°. The observed right ascensions and declinations derived from equations (B7) are in the epoch of the observations and the equator of the reference catalog. To avoid unnecessarily large allowances for the rotation and translation that precession will produce in each region, the reference catalog's equinox should be within the domain of the observational epochs.

The third step is the rigorous transformation of the reduced observations into the standard coordinate frame:

$$\xi = \frac{\cos\delta \sin(\alpha - \alpha_0)}{\sin\delta_0 \sin\delta + \cos\delta_0 \cos\delta \cos(\alpha - \alpha_0)} \tag{B8a}$$

and

$$\eta = \frac{\cos\delta_0 \sin\delta - \sin\delta_0 \cos\delta \cos(\alpha - \alpha_0)}{\sin\delta_0 \sin\delta + \cos\delta_0 \cos\delta \cos(\alpha - \alpha_0)} \tag{B8b}$$

where ξ and η should be multiplied by 206,264.806247 for conversion to arcseconds and (α_0, δ_0) is the tangent point that is adjusted on each pass to lie as near the center of the stars' apparent trajectory as possible. The η axis contains the projection of the great circle passing through the tangent point and the celestial poles; the ξ axis is perpendicular to the η axis at the tangent point, increasing with increasing right ascension.

The fourth and final transformation brings the axes of analysis in line with the original axes of observation. For the first ISI, we have:

$$F = \xi_F \cos\omega_F + \eta_F \sin\omega_F \tag{B9a}$$

and, for the second ISI, we have:

$$S = \xi_S \cos \omega_S + \eta_S \sin \omega_S \tag{B9b}$$

where F and S denote the derivation of respective sets of ξ and η, and in each case ω is derived from equation (B1) where $\omega = 90° - \theta$.

We now have the reduced x measurements of the two ISI's within the same coordinate system. However, we will not make use of this fact until the final stages of the current iteration. Instead we analyze the apparent star plotted against the F axis, which can be modeled:

$$F = F_0 + \mu_F t + \pi f_F + \frac{\mu'_F}{2} t^2 \tag{B10}$$

with a similar expression for the s axis, where t is the time since the star passed the tangent point, μ_F is the proper motion in the F coordinate, π is the trigonometric parallax, and μ'_F is the annual rate of change of μ_F (ref. 74). The factor f_F is derived from the usual parallax factors:

$$f_F = f_\alpha \cos \omega_F + f_\delta \sin \omega_F \tag{B11}$$

with a similar expression for f_S.

For purposes of this discussion, we will ignore a number of possible statistical refinements that present themselves at this point. The axes of observation are only perpendicular at one point observable by the Orion system and there are terms in the above and following equations of condition that would benefit from a unified reduction of all parameters of motion. But we will leave such things to those who may follow.

When the residuals of equation (B10) show an apparently non-random variation, two new terms may be added to the equation, namely:

$$F = F_0 + \mu_F t + \pi f_F + \frac{\mu_F'}{2} t^2 + Ix_F + Jy_F$$

or

$$S = S_0 + \mu_S t + \pi f_S + \frac{\mu_S'}{2} t^2 + Kx_S + Ly_S \qquad \text{(B12)}$$

for the δ coordinate. The terms x_F, y_F, x_S, and y_S are elliptic rectangular coordinates rotated into the respective coordinate systems, while I, J, K, and L are similar to the usual Thiel-Innes constants (see ref. 5).

The ecliptic coordinates are computed from the dynamical elements of the orbit: the period, time of periastron passage, and eccentricity. These are first estimated from the general form of the residuals to equation (B10) and then improved by a series of iterative adjustments to (B12) until the lowest sum of the squares of the residuals is attained (ref. 75). Equations (B10) and (B12) can take on more complicated forms for multiple-body systems, but these will be relatively infrequent.

An additional form of these equations is useful in attaching a statistical significance to any proposed perturbation:

$$F = F_0 + \mu_F t + \pi f_F + \frac{\mu_F'}{2} t^2 + \rho Q_F \qquad \text{(B13)}$$

with a similar expression in S. The term Q_F is computed using all that is known about the proposed orbit, while ρ, scaled to itself, should be identically 1 on each coordinate. Its variation from this value on the two axes gives a check on the agreement between the two ISI's while the ratio to its formal standard error gives an immediate indication of the statistical significance of the observed variations from the linear motion defined by equation (B10).

Presumably, such a test would be automatically run for each target star, allowing the investigator to attach a confidence level to

all nondetections as well as to suspected detections. The formal error computed for ρ in this manner will prove invaluable as a statistical tool for estimating the frequency of planetary systems. As stated elsewhere, the near-circular orbit of a star around the center of gravity it shares with its major planet will present a deflection of the star's trajectory that is evident in any orientation of the orbit. Detection depends only on the angular size of the orbit and the precision of the astrometric observations.

The user of the Orion system suffers some disadvantage because his astrometric precision is a function of declination and position angle. In more than two-thirds of the regions accessible to the system, the F and S axes are not even nearly perpendicular, but intersect at an angle of approximately 60°. As a result, the instrument has less than 60 percent of the sensitivity to orbits oriented in the north-south direction as to orbits oriented east-west. This, in turn, substantially affects the certainty with which the investigator will be able to say that a certain star does not have planets within given ranges of period and above a particular mass. In other words, the minimum detectable mass is not as small as the measuring precision suggests.

Once the form of equations (B10) and (B12) is determined and coefficients calculated in each coordinate, the output of the two ISI's may be finally combined to form improved predicted positions for the stars observed. Where the angles defined for the two ISI's are θ_F and θ_S, we have:

$$\left.\begin{aligned} \xi &= \frac{F/\cos\theta_S + S/\cos\theta_F}{\tan\theta_F + \tan\theta_S} \\ \text{and} \qquad \eta &= \frac{F\tan\theta_2/\cos\theta_1 - S\tan\theta_1/\cos\theta_2}{\tan\theta_F - \tan\theta_S} \end{aligned}\right\} \qquad \text{(B14)}$$

with F and S as defined previously. Hence, equations (B14) contain time-dependent functions.

Precession does present some difficulties. The slow change in the apparent intersection of the projected axis of Earth's rotation and the celestial sphere causes a variation in the angle at which a star

approaches the two vertical circles. This, in turn, changes the orientation of the two axes of measurement with respect to the trajectory of the star. Thus the one-dimensional observations are obtained in a rotating reference frame. However, the period of precession is long (about 26,000 years) in comparison to that of a proposed observational series. Thus, to sufficient precision, the rotation terms of equations (B4) and (B5) allow the necessary compensation. The angles θ_F and θ_S here become the initially estimated angles derived from equation (B1), with the coefficients of equations (B4) and (B5) providing the alignment necessary to counter the effects of small amounts of variability. The iterative nature of the central overlap technique ensures that the estimated y positions (γ) used for these small compensations in the latter equations are of acceptable precision.

The standard coordinates derived from equations (B14) are expressed in the equatorial coordinate system using the rigorous transformations:

$$\left.\begin{aligned} \alpha &= \alpha_0 + \arctan[\xi/(\cos\delta_0 - \eta \sin\delta_0)] \\ \delta &= \arcsin[(\sin\delta_0 + \eta\cos\delta_0)/(\xi^2 + \eta^2 + 1)^{1/2}] \end{aligned}\right\} \quad \text{(B15)}$$

where α and δ are in radians.

We have now completed the first full iteration. The values predicted by equations (B15) will have a much higher internal precision than those from the star catalog or astrographic photographs. But the new positions will retain the scale, orientation, mean position, and mean motion of the catalog positions. They will also retain the errors in these initial quantities, but these will have little effect on the information as seen. Many potential second- or higher-order spurious effects in the catalog positions will be smoothed out or completely eliminated by the overlap technique.

With careful attention to the detail of equations (B4), (B10), and (B12), we will note a drop in the formal error of their parameters during each of the first few iterations through equations (B2) to (B15). After three or four iterations, no parameter will change its value from one iteration to the next by more than 10 percent of its formal error. At this point, the iterative version of the central overlap technique is considered to have converged.

APPENDIX C

SONINE APODIZATION USING MASKS

Suppose we wish to simulate the Sonine distribution whose intensity transmission is

$$T_\mu(r) = \left(1 - \frac{r^2}{a^2}\right)^\mu \tag{C1}$$

The amplitude distribution for this case is

$$K_\mu(r) = \left(1 - \frac{r^2}{a^2}\right)^{\mu/2} \tag{C2}$$

and the equivalent one-dimensional distribution is

$$\begin{aligned} F_\mu(x) &= \int K_\mu(x^2 + y^2)dy \\ &= \sqrt{\pi a}\,\frac{\Gamma[(\mu+2)/2]}{\Gamma[(\mu+3)/2]}\left(1 - \frac{x^2}{a^2}\right)^{(\mu+1)/2} \end{aligned} \tag{C3}$$

Ignoring constant multipliers, we have for the mask shape:

$$f_\mu(x) = \pm a\left(1 - \frac{x^2}{a^2}\right)^{(\mu+1)/2} \tag{C4}$$

Figure 63 shows the masks for $\mu = 0, 1, 2, 3, 4$, and 5.

These masks will, of course, have only "good" directions parallel to the x axis. However, in these directions the performance should be as good as obtainable with a variable density mask in all

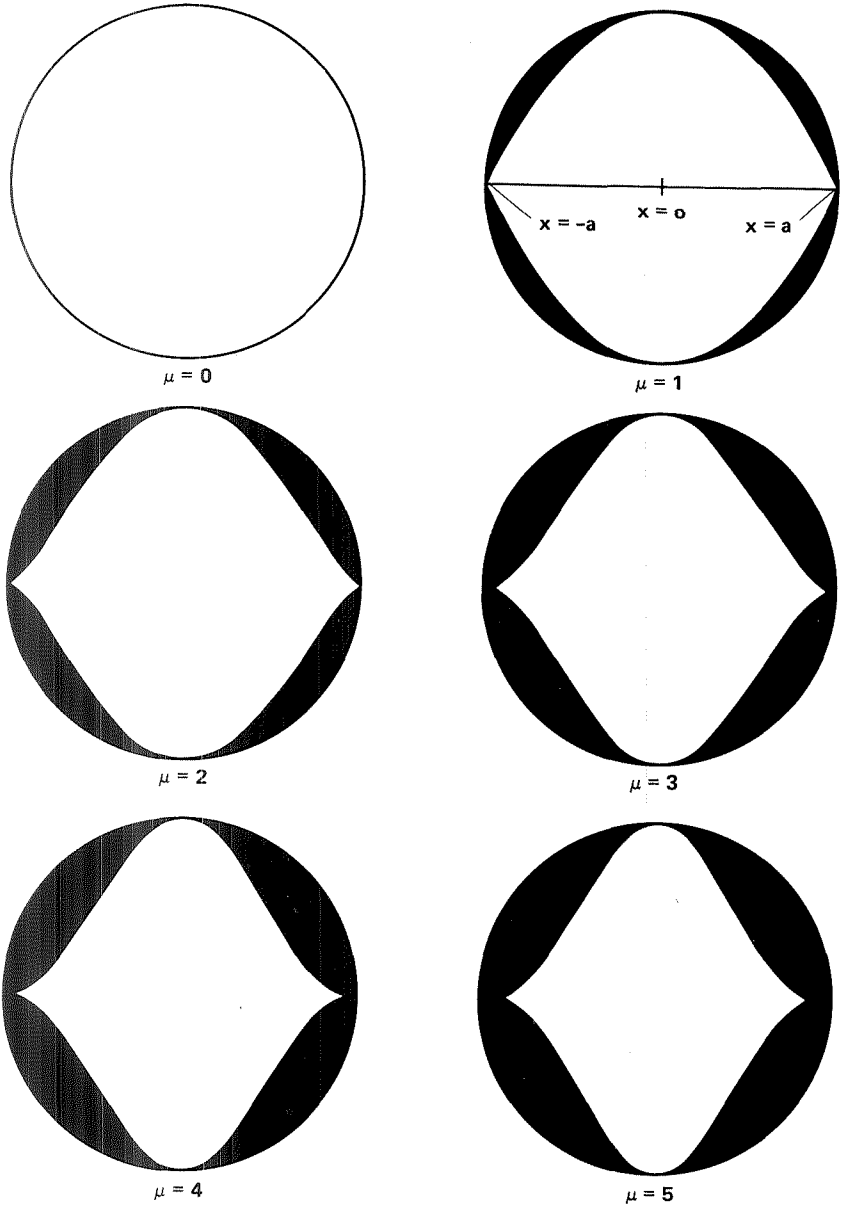


Figure 63.– *Schematic representation of the form of the one-dimensional Sonine masks used in the Orion apodization test; μ = 0 corresponds to a clear aperture; increasing μ values block out progressively more light. A complementary Sonine mask for a given μ value is one that is opaque where the corresponding Sonine mask transmits and transmitting where the Sonine mask is opaque.*

directions. This should allow an assessment of the light scattered by figuring errors and other imperfections.

APPENDIX D

DIRECT IMAGING OF EXTRASOLAR PLANETS WITH A SPACE TELESCOPE

If planetary systems are common in our galaxy, direct photography of these extrasolar planets would provide convincing proof of their existence. However, detecting the images of extrasolar planets presents a major technical problem: the extremely faint planetary image must be detected in the presence of background light, diffracted and scattered from the much brighter image of the planet's star. For example, if we were trying to detect Jupiter from a distance of 10 pc, Jupiter would be about 3×10^8 times fainter than the Sun and their separation would be at most 0.5 arcsec. Considerable reduction of the stellar background light is achieved with an occulting edge, used in conjunction with a telescope in space, as described by Spitzer (ref. 76). A practical realization of the occulting edge required by Spitzer's scheme would be a "black limb" of the Moon (i.e., a limb not illuminated by the Sun or Earth). Furthermore, the use of the lunar limb alleviates the critical alignment problems that would be encountered for a man-made occulting edge, necessarily of smaller size. The proposed 2.4-m space-located telescope (ref. 77) and its area photometer should be capable of detecting the Jupiter-Sun system at a distance of 10 pc. For this hypothetical case, the detection time would be short, requiring less than 20 min to detect Jupiter.

Starlight Attenuation and Signal-to-Noise Ratio

The relative intensity of the planetary image and background starlight in the focal plane of the telescope will be determined by the

starlight attenuation that can be produced by the occulting edge, in addition to the attenuation produced by the telescope itself. The situation is illustrated in figure 64, with the planet and its star at the left. The space telescope views the planet at an angle θ_2 above the horizon provided by the black lunar limb, while the star is at an angle θ_1 below this horizon. The diameter of the telescope aperture is d and it lies at a distance ℓ from the Moon. Since the telescope is within the geometrical shadow of the lunar limb, the intensity of starlight entering the telescope is greatly reduced from its unocculted value.

Here we shall consider the telescope to be an ideal optical system and neglect the effects of scattered light. The attenuation factor γ is defined as the amount of background starlight within the Airy disk containing the planet image divided by the peak intensity of the stellar image that would be obtained without an occulting disk. Modifying the equations given by Spitzer (ref. 76) to include the angles θ_1 and θ_2 as independent variables, we can write the attenuation factor γ as

$$\gamma = \underbrace{\left[\frac{1}{4\pi^2\theta_1^{\,2}}\left(\frac{\lambda}{\ell}\right)\right]}_{\text{attenuation by occulting edge}}\underbrace{\left[\frac{1}{0.23}\times\frac{1}{2\pi^4\theta_2^{\,3}}\left(\frac{\lambda}{d}\right)^3\right]}_{\text{attenuation by telescope}} = 5.7\times10^{-4}\frac{\lambda^4}{\theta_1^{\,2}\theta_2^{\,3}d^3\ell} \qquad \text{(D1)}$$

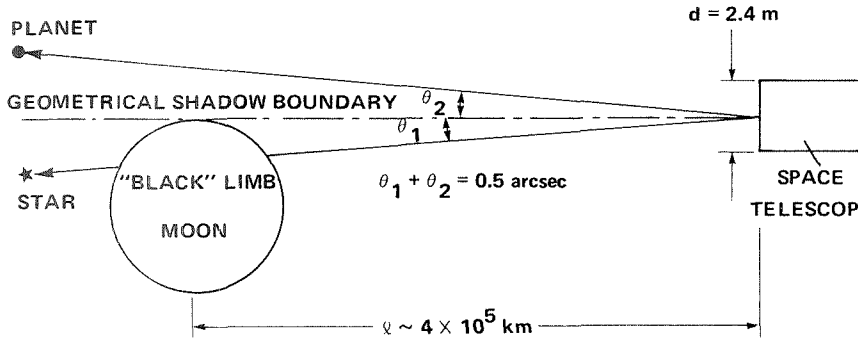


Figure 64.– *Schematic representation of the relative positioning of a space telescope and the Moon when the black limb of the Moon is used as an occulting edge.*

where 0.23 is the average intensity within the first dark ring of a diffraction pattern whose peak intensity is 1.0. Although the planet is at an angle $\theta_1 + \theta_2$ from the star, the separation will appear to be only θ_2 at the telescope because the light received from the star has been diffracted around the limb. In equation (D1), γ, d, and ℓ should be expressed in the same units, and θ_1 and θ_2 are in radians. We see that the factor γ has a strong dependence on the wavelength used, the telescope aperture, and the angular separation of the planet and its star.

For the Jupiter detection example, we let $\lambda = 5000$ Å, $\ell = 4\times10^5$ km, $a = 2.4$ m, and $\theta_1 = \theta_2 = 0.25$ arcsec $= 1.25\times10^{-6}$ rad. Then $\gamma = 2.5\times10^{-9}$, consisting of a 2.2×10^{-5} attenuation from the occulting edge and a 1.13×10^{-4} attenuation by the telescope itself. Since the Moon is dark and rough, specular reflections should not be a problem and the attenuation factor for the occulting edge should be a realistic estimate. However, real telescope optics would not be strictly diffraction-limited (as was assumed to write eq. (D1)) and scattered light within the telescope would raise the background light level. Apodization could be used to increase the attenuation, and here we shall assume that an attenuation of 10^{-4} could be achieved by the telescope.

Knowing the attenuation factor γ, we can now calculate the signal-to-noise ratio for detecting Jupiter at a distance of 10 pc. The photon-counting area photometer described by Laurance (ref. 78) would be used at the focus of the telescope. We let f_p, f_*, and f_S equal the fluxes from the planet, star, and the sky, respectively, in photons cm^{-2} Å^{-1} sec^{-1} within the Airy disk containing the planet's image. The quantity f_n is the detector noise, expressed as an equivalent photon flux. We assume that the photon counts recorded by the detector obey Poisson statistics. The signal-to-noise ratio (S/N) is given by

$$\frac{S}{N} = \frac{f_p(qA\Delta\lambda T)^{1/2}}{(f_p + \gamma f_* + f_S + f_n)^{1/2}} \tag{D2}$$

The quantity q is the photon detection efficiency of the telescope optics, filter, and detector system; A is the area of the telescope; $\Delta\lambda$ is the wavelength passband of the filter; and T is integration time.

From equation (D2), we can compute S/N for detecting Jupiter at a distance of 10 pc. For these conditions, the visual magnitude (m_v) of the Sun would be +4.8. Fully illuminated by the Sun, Jupiter would be 21 magnitudes fainter, but, at their maximum apparent separation, Jupiter would appear only half-illuminated. Allowing one magnitude for this effect, we obtain, for Jupiter, $m_v = 26.8$. Hence $f_p = 1.9 \times 10^{-8}$ photons cm^{-2} Å^{-1} sec^{-1} and $\gamma f_* = 3.0 \times 10^{-8}$ photons cm^{-2} Å^{-1} sec^{-1}. The sky flux f_S should be less than 10^{-8} within the Airy disk of the planet image, and we assume the detector noise f_n is negligible, too.

For our example we let $T = 1000$ sec, $\Delta\lambda = 1000$ Å, $A = 4.5 \times 10^4$ cm^2, and $q = 0.25$; this yields S/N ~ 8 from equation (D2). Hence, Jupiter could be detected in less than 20 min with a good S/N. For a telescope with an aperture d smaller than 2.4 m, the integration time required to achieve the same S/N increases considerably. For smaller values of d, the main source of noise will most likely be scattered light from the star, so in equation (D2) we can make the approximation $\gamma f_* >> f_p + f_S + f_n$. If all other factors are constant, the integration time $T \propto d^{-5}$. Hence about 80,000 sec (1 day) of integration time would be required if a 1-m telescope instead of a 2.4-m telescope were used for the above example.

Another consideration for the required integration time is the distance of the planet from its star. Again, we make the approximation $\gamma f_* >> f_p + f_S + f_n$ and let $\theta_1 + \theta_2 = a/D$, where a is the semi-major axis of the planet's orbit and D is the distance to the star. Letting $\theta_1 = \theta_2$, we find the integration time $T \propto a^{-1}$. Hence, the required integration time will be less for planets farther away from their star, but the dependence is not nearly so strong as for the other variables.

After a faint image is detected near a star, several criteria can be applied to confirm that it is a planet. First, its intensity and distance from the star must lie within acceptable limits. Then, from several months of repeated measurements, the proper motion of the suspected planet could be checked against that of the star. If they agree, this would be strong evidence that the body is indeed a planet. Depending on S/N, it might be possible to measure a crude spectrum of the planet and search for the deep methane absorption bands in the far red, characteristic of Jupiter and the other large planets of our solar system.

If the planet's orbit has a favorable orientation and S/N is sufficient, we can determine the inclination, eccentricity, and semimajor axis of the orbit by measuring the position of the planet relative to the star through one orbital period (12 years for Jupiter, 30 years for Saturn). The solution for these orbital parameters (and the mass of the star) would use the same technique as for visual binary stars of known parallax (ref. 79).

Recommendations for Further Development

From the calculations above, we can be optimistic about detecting extrasolar planets with a 2.4-m telescope in space and its proposed state-of-the-art instrumentation. We note that 37 nearby stars are brighter than $m_v = +4.8$ and are closer than 10 pc. Hence, the detection of a Jupiter-type planet orbiting any of these stars would be no more difficult than our Jupiter-Sun example. For the brightest and nearest stars, S/N should be substantially greater than in our example.

Several technical problems must be investigated before we can be confident that a space telescope used with an occulting edge would achieve the desired sensitivity for the detection of extrasolar planets. An adequately sensitive photon-counting area photometer must be constructed and the calculated starlight attenuation factors (eq. (D1)) must be demonstrated to be achievable in practice.

Another important consideration is the orbit of the telescope. The near-Earth orbit planned for the space telescope will not be suitable because (1) the lunar aspect seen by the telescope will not be sufficiently free from earthshine, (2) the relative velocity of the telescope and the lunar limb will be too great to allow the appropriate alignment to be maintained for the necessary integration times, and (3) most of the 37 stars that are within 10 pc and are brighter than $m_v = +4.8$ would not be occulted by the Moon.

To overcome those difficulties, an orbit for the telescope comparable in size to that of the Moon is needed, and the telescope orbit must be changed from time to time – perhaps by solar sail – to achieve the proper alignments with the lunar limb. Choosing an optimum strategy for the telescope orbit and its required changes presents an interesting, difficult, and important problem.

If the S/N estimates presented here can be substantiated by laboratory tests, then we should determine what would be required to raise the orbit of the space telescope, at some point during its lifetime, in order to search for extrasolar planets. Meanwhile, the necessary experience for operating a telescope in space with the black limb of the Moon as an occulting disk could be gained by operating a smaller telescope in the required orbit configuration. In addition to preparing for planet detection, such a mission would seem capable of many important astronomical observations. For example, a 1-m telescope equipped with a simple optical photometer in high orbit could be used to record lunar occultations of virtually any object in the sky. The additional advantages of the black lunar limb and much slower occultation rates would permit the measurement of many more stellar diameters (ref. 80), binary star separations, and the angular diameters of quasars.

In addition, the large parallax allowed by a high orbit would permit many more stellar occultations by solar system objects to be observed. Some of the exciting work that could be accomplished is: (1) further investigation of the newly discovered rings of Uranus (refs. 81 and 82); (2) search for rings around Neptune and Jupiter; (3) investigation of the optical depth of Saturn's rings with resolution of a few kilometers; (4) determination of diameters of Pluto, satellites, and asteroids; (5) further investigation of the tidal waves in the Martian atmosphere (ref. 83); and (6) acquisition of many temperature, pressure, and number density profiles of the upper atmospheres of all planets, including Earth (ref. 84).

Such a dedicated occultation telescope could have an extremely productive lifetime in the course of perfecting techniques that would ultimately be used for detecting extrasolar planets.

REFERENCES

1. van de Kamp, Peter: Astrometric Study of Barnard's Star from Plates Taken with the 24-inch Sproul Refractor. Astron. J., vol. 68, no. 7, Sept. 1963, pp. 515–521.

2. van de Kamp, Peter: Alternate Dynamical Analysis of Barnard's Star. Astron. J., vol. 74, no. 6, Aug. 1969, pp. 757–759.

3. Gatewood, George: An Astrometric Study of Kuiper 84 (BD + 4°3562). Astron. J., vol. 78, no. 8, Oct. 1973, pp. 777–779.

4. Hershey, John L.: Astrometric Analysis of the Field of AC + 65°6955 from Plates Taken with the Sproul 24-inch Refractor. Astron. J., vol. 78, no. 5, June 1973, pp. 421–425.

5. van de Kamp, P.: Principles of Astrometry, W. H. Freeman & Co., 1967.

6. Reiger, S. H.: Starlight Scintillation and Atmospheric Turbulence. Astron. J., vol. 68, no. 6, Aug. 1963, pp. 395–406.

7. Young, Andrew T.: Photometric Error Analysis. VIII. The Temporal Power Spectrum of Scintillation. Appl. Optics, vol. 8, no. 5, May 1969, pp. 869–885.

8. Gatewood, G.: On the Astrometric Detection of Neighboring Planetary Systems. Icarus, vol. 27, no. 1, 1976, pp. 1–12.

9. Born, Max; and Wolf, Emil: Principles of Optics. Electromagnetic Theory of Propagation, Interference and Diffraction of Light. Third ed., Pergamon Press, Oxford, New York, 1965, p. 218.

10. Strand, K. Aa.: 61-Inch Astrometric Reflector System, with a List of References. Publications of the U.S. Naval Observatory, Second Series, vol. 20, part I, 1971.

11. Kraushaar, R.: A Diffraction Grating Interferometer. J. Opt. Soc. Am., vol. 40, no. 7, July 1950, pp. 480–481.

12. Meyer-Arendt, Jurgen; and Miner, Ellis D., Jr.: Angular Measurements by Means of a Ronchi Ruling. Appl. Optics, vol. 2, no. 1, Jan. 1963, pp. 77–78.

13. Yokozeki, S.; and Suzuki, T.: Shearing Interferometer Using the Grating as the Beam Splitter. Appl. Optics, vol. 10, no. 7, July 1971, pp. 1575–1580.

14. Theocaris, P. S.; and Liakopoulos, C.: Alignment, Orientation, and Range Finding by Diffraction Gratings. Appl. Optics, vol. 12, no. 10, Oct. 1973, pp. 2288–2297.

15. Hong, Yie-Ming: Tilted-Reference-Beam Lateral Shearing Interferometer Utilizing Carrier-Frequency Photography. Appl. Optics, vol. 14, no. 3, March 1975, pp. 614–617.

16. Yokozeki, Shunsuke; and Ohnishi, Kazuo: Spherical Aberration Measurement with Shearing Interferometer Using Fourier Imaging and Moiré Method. Appl. Optics. vol. 14, no. 3, March 1975, pp. 623–627.

17. Patorski, Krzyzto; Yokozeki, Shunsuke; and Suzuki, Tatsuro: Collimation Test by Double Grating Shearing Interferometer. Appl. Optics, vol. 15, no. 5, May 1976, pp. 1234–1240.

18. Talbot, H.: Facts Relating to Optical Science 4. Phil. Mag., vol. 9, 1836, pp. 401–409.

19. Rayleigh, Lord: On Copying Diffraction Gratings and Some Phenomena Connected Therewith. Phil. Mag., vol. 11, 1881, pp. 196–205.

20. Malacara, D.; and Cornejo, A.: The Talbot Effect in the Ronchi Test. Boletin Inst. Tonantzintla, vol. 1, no. 3, Dec. 1974, pp. 193–196.

21. Ronchi, Vasco: Forty Years of History of a Grating Interferometer. Appl. Optics, vol. 3, no. 4, April 1964, pp. 437–451.

22. Miller, R. H.: Measurement of Stellar Diameters. Science, vol. 153, no. 3736, Aug. 5, 1966, pp. 581–587.

23. Jacquinot, P.: Apodisation. Progress in Optics, Vol. III, E. Wolf, ed., North-Holland Publishing Co., Amsterdam, 1964, pp. 31–186.

24. Lindenblad, Irving W.: Relative Photographic Positions and Magnitude Difference of the Components of Sirius. Astron. J., vol. 75, no. 7, Sept. 1970, pp. 841–847.

25. van Albada, G. B.: Contribution Bosscha Observatory, 1958a, no. 6, p. 3.

26. Tuvikene, L. M.: On Apodizating Diaphragms and Diaphragms Increasing Resolving Power. Optics and Spectroscopy, vol. 10, no. 2, Feb. 1961, pp. 144–145.

27. Barlow, Boris V.: The Astronomical Telescope. Springer-Verlag, New York, 1975, pp. 25–26.

28. KenKnight, C. E.: Methods of Detecting Extrasolar Planets. 1. Imaging. Icarus, vol. 30, no. 2, Feb. 1977, pp. 422–433.

29. Born, Max; and Wolf, Emil: Principles of Optics. Ch. 8. Fourth ed., Pergamon Press, Oxford, New York, 1970, p. 415, pp. 275–276, 333–338, and 312–315.

30. Ditchburn, R. W. (F.R.S.): Light. Vol. 1, Second ed., Interscience Publishers, New York, 1963, pp. 200–202.

31. Robinson, L. C.: Physical Principle of Far-Infrared Radiation. Methods of Experimental Physics. Vol. 10, ch. 4. Academic Press, New York, 1973.

32. Simpson, J. P.; and Witteborn, F. C.: Effect of the Shuttle Contaminant Environment on a Sensitive Infrared Telescope. Appl. Optics, vol. 16, no. 8, Aug. 1977, pp. 2051–2073.

33. Putley, E. H.: Solid State Devices for Infra-red Detection. J. Sci. Instrum., vol. 43, no. 12, Dec. 1966, pp. 857–868.

34. Putley, E. H.: Modern Infrared Detectors. Physics in Technology, vol. 4, no. 3, Dec. 1973, pp. 202–222.

35. Thiessen, G.; and Broglia, P.: A Polarization Effect in (Vacuum) Evaporated Aluminum Films in the Case of Normal (Unpolarized) Light Incidence. Z. Astrophys., vol. 48, no. 2, pp. 81–87, 1959 (in German).

36. Maksutov, Dmitry Dmitrievich: Astronomicheskaya Optika (Astronomical Optics). Moscow, 1946.

37. Gascoigne, S. C. B.: Some Recent Advances in the Optics of Large Telescopes. Royal Astron. Soc., Quarterly J., vol. 9, no. 2, June 1968, pp. 98–115.

38. Crawford, D. L.; Meinel, A. B.; and Stockton, Martha W., eds.: A Symposium on Support and Testing of Large Astronomical Mirrors, Kitt Peak National Observatory and University of Arizona, Tech. Rept. 30, University of Arizona, July 1968.

39. Hoag, A. A.; Priser, J. B.; Riddle, R. K.; and Christy, J. W.: Installation, Tests, and Initial Performance of the 61-Inch Astrometric Reflector. Publications of the United States Naval Observatory, Second Series, Vol. XX, Part II, 1967.

40. Fuller, Dudley, D.: Theory and Practice of Lubrication for Engineers. John Wiley & Sons, New York, 1956.

41. West, Richard M., ed.: Proceedings, ESO/CERN Conference on Large Telescope Design, Geneva, March 1–5, 1971, Presses du courier de Geneve, 1971, p. 405.

42. Groeneveld, D. G. S.: Considerations in the Design of Primary Worm-Gear Drives for Astronomical Telescopes. Astron. Soc. Australia, Proceedings, vol. 1, no. 5, March 1969, pp. 245–247.

43. Betz, H. D.: An Asymmetry Method for High Precision Alignment with Laser Light. Appl. Optics. vol. 8, no. 5, May 1969, pp. 1007–1013.

44. Merceret, Francis J.: Airborne Hot-Film Measurements of the Small-Scale Structure of Atmospheric Turbulence During GATE. J. Atmos. Sci., vol. 33, 1976, pp. 1739–1746.

45. Lawrence, R. S.; Ochs, G. R.; and Clifford, S. F.: Measurements of Atmospheric Turbulence Relevant to Optical Propagation. J. Opt. Soc. Am., vol. 60, 1970, pp. 826–830.

46. Bufton, Jack L.; Minott, Peter O.; Fitzmaurice, Michael W.; and Titterton, Paul J.: Measurements of Turbulence Profiles in the Troposphere. J. Opt. Soc. Am., vol. 62, 1972, p. 1068.

47. Bufton, Jack L.: Correlation of Microthermal Turbulence Data with Meteorological Soundings in the Troposphere. J. Atmos. Sci., vol. 30, 1973, pp. 83–87.

48. Bufton, Jack L.: Comparison of Vertical Profile Turbulence Structure with Stellar Observations. Appl. Optics, vol. 12, 1973, pp. 1785–1793.

49. Taylor, G. I.: Statistical Theory of Turbulence, I. Proc. Roy. Soc. (London), vol. A151, 1935, p. 421.

50. Corrsin, Stanley: On the Spectrum of Isotropic Temperature Fluctuations in an Isotropic Turbulence. J. Appl. Phys., vol. 22, 1951, pp. 469–473.

51. Molyneux, John E.: Analyses of "Dishonest" Methods in the Theory of Wave Propagation in a Random Medium. J. Opt. Soc. Am., vol. 58, 1968, pp. 951–957.

52. von Kármán, T.: Progress in Statistical Theory of Turbulence. Proc. Natl. Acad. Sci. (USA), vol. 34, 1948, pp. 530–539.

53. Kolmogorov, A. N.: Compt. rend. (Doklady) de l'Acad. Sci. URSS, vol. 30, 1941, p. 301.

54. Kolmogorov, A. N.: Compt. rend. (Doklady) de l'Acad. Sci. URSS, vol. 32, 1941, p. 16.

55. Onsager, Lars: The Distribution of Energy in Turbulence. Phys. Rev., vol. 68, 1945, p. 286.

56. von Weizsacker, C. F.: Spectrum of Turbulence at Large Reynolds Numbers. Z. Phys., vol. 124, 1948, pp. 614–627.

57. Pasqualetti, F.; Ronchi, L.; and Vanni, R.: Direct Measurements of the Model of the Atmospheric Turbulence. Appl. Optics, vol. 15, 1976, pp. 1903–1904.

58. Fried, D. L.: Diffusion Analyses for the Propagation of Mutual Coherence. J. Opt. Soc. Am., vol. 58, 1968, pp. 961–969.

59. Cook, Richard J.: Beam Wander in a Turbulent Medium: An Application of Ehrenfest's Theorem. J. Opt. Soc. Am., vol. 65, 1975, pp. 942–948.

60. Hufnagel, R. E.; and Stanley, N. R.: Modulation Transfer Function Associated with Image Transmission Through Turbulent Media. J. Opt. Soc. Am., vol. 54, 1964, pp. 52–61.

61. Lin, C. C.: Note on Law of Decay of Isotropic Turbulence. Proc. Nat. Acad. Sci. (USA), vol. 34, 1948, pp. 540–543.

62. Tatarski, Valerian I.: Wave Propagation in a Turbulent Medium. McGraw-Hill Book Co., New York, 1961.

63. Erdélyi, Arthur, ed.: Tables of Integral Transforms, Vol. 2. McGraw-Hill Book Co., New York, 1954.

64. Bouricius, G. M. B.; and Clifford, S. F.: Experimental Study of Atmospherically Induced Phase Fluctuations in an Optical Signal. J. Opt. Soc. Am., vol. 60, no. 11, Nov. 1970, pp. 1484–1489.

65. Fried, D. L.: Statistics of a Geometric Representation of a Wavefront Distortion. J. Opt. Soc. Am., vol. 55, no. 11, Nov. 1965, pp. 1427–1435.

66. Wesely, M. L.; and Derzko, Z. I.: Atmospheric Turbulence Parameters from Visual Resolution. Appl. Optics, vol. 14, no. 4, April 1975, pp. 847–853.

67. Fried, D. L.: Optical Resolution Through a Randomly Inhomogeneous Medium for Very Long and Very Short Exposures. J. Opt. Soc. Am., vol. 56, no. 10, Oct. 1966, pp. 1372–1379.

68. Breckenridge, J. B.: Measurement of the Amplitude of Phase Excursions in the Earth's Atmosphere. J. Opt. Soc. Am., vol. 66, no. 2, Feb. 1976, pp. 143–147.

69. KenKnight, C. E.: Autocorrelation Methods to Obtain Diffraction-Limited Resolution with Large Telescopes. Astrophys. J., vol. 176, no. 1, pt. 2, Aug. 15, 1972, pp. L43–L45.

70. Chandrasekhar, S.: A Statistical Basis for the Theory of Stellar Scintillation. Monthly Notices Roy. Astron. Soc. (London), vol. 112, no. 5, 1952, pp. 475–483.

71. Eichhorn, H.: Zür Bestimmung von photographischen Sternpositionen und Eigenbewungen. Astron. Nachr., vol. 285, 1960, p. 233.

72. Schlesinger, F.: Irregularities in Atmospheric Refraction. Publications of the Allegheny Observatory, vol. 4, 1916, p. 1.

73. Hudson, C.: Irregularities in Refraction. Publications of the Allegheny Observatory, vol. 6, 1929, p. 1.

74. Gatewood, George; and Russell, Jane: Astrometric Determination of the Gravitational Redshift of van Maanen 2 (EG 5). Astron. J., vol. 79, no. 7, July 1974, pp. 815–818.

75. Gatewood, George: An Astrometric Study of Lalande 21185. Astron. J., vol. 79, no. 1, Jan. 1974, pp. 52–53.

76. Spitzer, Lyman, Jr.: The Beginnings and Future of Space Astronomy. Am. Sci., vol. 50, no. 3, Sept. 1962, pp. 473–484.

77. The Space Telescope. NASA SP-392, 1976.

78. Laurance, R. J.: The European Space Agency Study of Photon Counting Imaging for the Space Telescope. The Space Telescope, NASA SP-392, 1976, pp. 97–105.

79. Binnendijk, Lendert: Properties of Double Stars. W. J. Mackay & Co., Ltd., Chatham, 1960, p. 56.

80. Barnes III, T. G.; and Evans, D. S.: Stellar Angular Diameters and Visual Surface Brightness. I. Late Spectral Types. Monthly Notices Roy. Astron. Soc., vol. 174, no. 3, March 1976, pp. 489–593.

81. Elliot, J. L.; French R. G.; Dunham, E. W.; Gierasch, P. J.; Veverka, J.; Church, C.; and Sagan, Carl: Occultation of ϵ Geminorum by Mars: Evidence for Atmospheric Tides? Science, vol. 195, no. 4276, Jan. 28, 1977, pp. 485–486.

82. Millis, R. L.; Wasserman, L. H.; and Birch, P. V.: Detection of Rings Around Uranus. Nature, vol. 267, no. 5609, May 26, 1977, pp. 330–331.

83. Elliot, J. L.; Dunham, E.; and Mink, D.: The Rings of Uranus. Nature, vol. 267, no. 5609, May 26, 1977, pp. 328–330.

84. Elliot, J. L.; and Veverka, J.: Stellar Occultation Spikes as Probes of Atmospheric Structure and Composition. Icarus, vol. 27, no. 3, March 1976, pp. 359–386.

Participants in the 1976 NASA/ASEE-Stanford Summer Faculty Workshop in Engineering Systems Design
June 14 – August 20, 1976

FACULTY FELLOWS

Shu-Bing Chan, California State College, Pomona
Shyming Chang, University of Alabama
Artice Davis, San Jose State University
Kuan-Chen Fu, University of Toledo
James Gilfert, Ohio University
Francis Griggs, Merrimack College
Marvin Heifetz, Milwaukee School of Engineering
Teruo Ishihara, Rose-Hulman Institute of Technology
Charles KenKnight, University of Arizona
Eugene Kosso, University of Nevada, Reno
John Mann, University of Illinois, Urbana
Krzysztof Serkowski, University of Arizona
Syed Shariq, Oklahoma State University
Orestes Stavroudis, University of Arizona
Patrick Takahashi, University of Hawaii, Manoa
Raymond Vito, Georgia Institute of Technology
Frissell Walker, The Federal City College, Washington, D.C.
Ji Wang, San Jose State University
Robert Waters, University of Missouri, Rolla
Theodore Zsutty, San Jose State University

PROJECT ADVISORS

Ronald N. Bracewell, Stanford University
George D. Gatewood, Allegheny Observatory, University of Pittsburgh

PROJECT DIRECTORS

David C. Black, NASA Ames Research Center
Robert Piziali, Stanford University

EDITOR'S NOTE

Our study has benefited greatly from advice and consultation with many people. Early in our activity, a lecture series (outlined below) was beneficial in acquainting the Fellows with the many facets of our problem:

1. A General Discussion of the Planetary Detection Problem and Methods for Detecting Other Planetary Systems – *D. C. Black, NASA Ames*

2. SETI and Planetary Detection – *H. Mark, Director, NASA Ames*

3. An Overview of the SETI Program – *J. Billingham, NASA Ames*

4. Possible Means of Detecting Other Planetary Systems at Infrared and Radio Frequencies – *T. Clark, NASA Goddard and University of Maryland*

5. Direct Imaging of Extrasolar Planets with an Apodized Space Telescope – *B. M. Oliver, Hewlett-Packard*

6. Astrometric Theory – *R. Harrington, U.S. Naval Observatory*

7. Long and Short Term Stability of Astrometric Telescopes – *K. A. Strand, Director, U.S. Naval Observatory*

8. Comments on Optical Requirements of an Astrometric Telescope – *C. KenKnight, University of Arizona*

9. Reduction of Astrometric Data – *G. Gatewood, Allegheny Observatory*

In addition, the group also heard lectures by D. Currie of the University of Maryland and R. Miller of the University of Chicago on the concept of a ground-based Michelson interferometer for

astrometric observations. Their contributions were invaluable in clarifying some of the concepts embodied in the imaging interferometer design concept developed here. We thank J. Elliot of Cornell University for his input to our study, most notably in the area of direct detection of extrasolar planets. A discussion by Dr. Elliot of one possible alternative to the techniques considered during Project Orion is presented in appendix D. Many people in local industry also kindly provided valuable advice. Those of us in the program acknowledge the support and contributions of all these people and others too numerous to mention.